

# Convergence of Imprecise Continuous-Time Markov Chains

*Jasper De Bock*

## Abstract

We study the limit behaviour of a generally non-linear ordinary differential equation whose solution is a superadditive generalisation of a stochastic matrix, and provide necessary and sufficient conditions for this solution to be ergodic, in the sense that it converges to an operator that, essentially, maps functions to constants. In the linear case, the solution of our differential equation is equal to the matrix exponential of an intensity matrix and can then be interpreted as the transition operator of a homogeneous continuous-time Markov chain. Similarly, in the generalised non-linear case that we consider, the solution can be interpreted as the lower transition operator of a specific set of non-homogeneous continuous-time Markov chains, called an imprecise continuous-time Markov chain. In this context, our main result provides a necessary and sufficient condition for such an imprecise continuous-time Markov chain to converge to a unique limiting distribution.

**Keywords:** Markov chain, continuous-time, imprecise, convergence, limiting distribution, ergodicity, matrix exponential, lower transition operator, lower transition rate operator.

## 1 Introduction

Consider a real-valued  $n \times n$  matrix  $Q$  and let  $T_t$  be a real-valued time-dependent  $n \times n$  matrix such that

$$\frac{d}{dt}T_t = QT_t \text{ for all } t \geq 0$$

and  $T_0 = I$ , with  $I$  the  $n$ -dimensional unit matrix. The unique solution of this differential equation is then well known to be given by the matrix exponential  $e^{Qt}$  of  $Q$ . If  $Q$  is furthermore an intensity matrix—has non-negative off-diagonal elements and rows that sum to zero—then  $T_t = e^{Qt}$  will be a stochastic matrix. In that case,  $T_t$  can be interpreted as the transition operator of a homogeneous continuous-time Markov chain. Indeed, if we identify  $\{1, \dots, n\}$  with the state space  $\mathcal{X}$  of such a Markov chain and let  $Q$  be its transition rate matrix, then for any two states  $x, y \in \mathcal{X}$ ,  $T_t(x, y)$  is the probability  $P(X_t = y | X_0 = x)$  of ending up in state  $y$  at time  $t$ , conditional on starting in state  $x$  at time zero.

Rather remarkably, for any transition rate matrix  $Q$ , the conditional probability  $P(X_t = y | X_0 = x)$  will always converge [1, Theorem II.10.1]. However, in general, this limiting value may depend on the initial state  $x$ . If this is not the case, that is, if there is a probability mass function  $P_\infty$  on  $\mathcal{X}$  such that

$$\lim_{t \rightarrow +\infty} P(X_t = y | X_0 = x) = P_\infty(y) \text{ for all } y \in \mathcal{X},$$

then the homogeneous continuous-time Markov chain under consideration—or, equivalently, the transition rate matrix  $Q$ —is said to have a unique limiting distribution  $P_\infty$ . From an applied point of view, the existence of such a limiting distribution is clearly important, because it implies that for large enough values of  $t$ , predicting the current value of  $X_t$  does not require any knowledge about its initial values. Hence, we are led the following question: what conditions does  $Q$  need to satisfy in order for  $P_\infty$  to exist? As it turns out, this question has an elegant answer: the required conditions are relatively easy, and are fully determined by the signs of the components of  $Q$ ; see for example Anderson [2].

Our main goal here is to answer a non-linear generalisation of this question, which includes the linear version that we have just discussed as a special case. Basically, the only difference is that the intensity matrix  $Q$  is replaced by a lower transition rate operator  $\underline{Q}$ , which is a non-linear—superadditive—generalisation of an intensity matrix. Much as in the original case, this lower transition rate operator gives rise to a corresponding lower transition operator  $\underline{T}_t$ , which is a non-linear—superadditive—generalisation of a stochastic matrix. For every real-valued function  $f$  on  $\mathcal{X}$ ,  $\underline{T}_t f$  is completely determined by the non-linear differential equation

$$\frac{d}{dt} \underline{T}_t f = \underline{Q} \underline{T}_t f \quad \text{for all } t \geq 0, \quad (1)$$

with boundary condition  $\underline{T}_0 f = f$  [3]. The aim of this paper is to study the properties of this operator  $\underline{T}_t$  and, in particular, its limit behaviour as  $t$  approaches infinity. Our main contribution—see Theorem 15—is a simple necessary and sufficient condition for  $\underline{Q}$  to be ergodic, in the sense that for all real-valued functions  $f$  on  $\mathcal{X}$ ,  $\lim_{t \rightarrow +\infty} \underline{T}_t f$  exists and is constant.

Our motivation for studying this property, and the reason for this papers title, is that  $\underline{T}_t f(x)$  can be interpreted as the conditional lower expectation  $\underline{\mathbb{E}}(f(X_t)|X_0 = x)$  of an imprecise continuous-time Markov chain, which, basically, is a set of continuous-time Markov chains whose possibly time-dependent transition rate matrix  $Q_t$  is partially specified, in the sense that all that we know about it is that it takes values in some given set of transition rate matrices  $\mathcal{Q}$ .<sup>1</sup> Indeed, as recently shown in References [4, 3], for the largest such set of Markov chains, and under relatively mild conditions on  $\mathcal{Q}$ ,<sup>2</sup> the tightest possible lower bound on the conditional expectation  $\mathbb{E}(f(X_t)|X_0 = x)$ —the conditional lower expectation  $\underline{\mathbb{E}}(f(X_t)|X_0 = x)$ —is equal to the solution  $\underline{T}_t f$  of the differential Equation (1), with  $\underline{Q}$  the lower envelope of  $\mathcal{Q}$ .

Therefore, if  $\lim_{t \rightarrow +\infty} \underline{T}_t f$  exists and is constant—if  $\underline{Q}$  is ergodic—this can be interpreted to mean that the limit value of the conditional lower expectation  $\underline{\mathbb{E}}(f(X_t)|X_0 = x)$  does not depend on the initial state  $x$ , or equivalently, that the imprecise continuous-time Markov chain under study has a unique limiting lower expectation operator  $\underline{\mathbb{E}}_\infty$ , in the sense that

$$\underline{\mathbb{E}}_\infty(f) = \lim_{t \rightarrow +\infty} \underline{\mathbb{E}}(f(X_t)|X_0 = x) \quad \text{for all } x \in \mathcal{X} \text{ and all real functions } f \text{ on } \mathcal{X}.$$

<sup>1</sup> In fact,  $\underline{T}_t f$  can also be interpreted as the conditional lower expectation  $\underline{\mathbb{E}}(f(X_t)|X_0 = x)$  that corresponds to an even larger set of stochastic processes, which, loosely speaking, is a set of (not necessarily Markovian) stochastic processes whose (possibly time- and history-dependent) transition rate matrix is not exactly specified, but is only known to take values in  $\mathcal{Q}$ ; see Reference [4] for more information.

<sup>2</sup> It should have separately specified rows, which means that every row has a set of possible candidate rows, and that the set of rate matrices  $\mathcal{Q}$  is constructed by combining these candidate rows in all possible ways, by selecting one element from each candidate set.

This interpretation turns our main result—a necessary and sufficient condition for  $\underline{Q}$  to be ergodic—into a practical tool: it provides a simple criterion for checking whether or not a given imprecise continuous-time Markov chain has a unique limiting lower expectation operator  $\mathbb{E}_\infty$ . In the special case where the lower transition rate operator  $\underline{Q}$  is actually a transition rate matrix  $Q$ , our notion of ergodicity coincides with the usual one and, in that case, our results can be used to check whether the continuous-time Markov chain that corresponds to  $Q$  has a unique limiting distribution  $P_\infty$ , whose expectation operator  $\mathbb{E}_\infty$  will then be equal to  $\mathbb{E}_\infty$ .

That being said, this paper does not adopt any specific interpretation, but takes a purely mathematical point of view. Our object of study here is the solution  $\underline{T}_t$  of the differential Equation (1), and our main result is a necessary and sufficient condition for  $\underline{Q}$  to be ergodic, in the sense that  $\underline{T}_t$  converges to a limit operator that maps functions to constants. As explained above, this result is directly applicable to—and inspired by—the theory of imprecise continuous-time Markov chains; more information about this field of study can be found in References [4, 3, 5, 6]. However, we think that our results should also be of interest to other fields whose aim it is to robustify the theory of continuous-time Markov chains, such as continuous-time Markov decision processes [7], continuous-time controlled Markov chains [8] and interval continuous-time Markov chains [9]. More generally, we believe that our ideas and results are relevant to any theory that studies—or requires—some kind of robust generalisation of the matrix exponential of an intensity matrix.

We end this introduction with a brief overview of the structure of this paper. After Section 2, in which we introduce some basic preliminary concepts, the rest of this paper is structured as follows.

We start in Section 3 by introducing the concept of a lower transition operator  $\underline{T}$ , which is a non-linear—superadditive—generalisation of a stochastic matrix; the operator  $\underline{T}_t$  that is studied in this paper is a special case. We provide a definition, explain the connection with coherent lower previsions [10, 11], and use this connection to establish a number of technical properties.

Section 4 then goes on to define ergodicity for lower transition operators, which is a discrete-time version of the notion of ergodicity that we study in this paper, and recalls that a lower transition operator  $\underline{T}$  will exhibit this type of ergodicity if and only if it is regularly absorbing [12]. We also introduce a new property, called being 1-step absorbing, and show that it is a sufficient condition for  $\underline{T}$  to be ergodic.

Next, in Section 5, we introduce the concept of a lower transition rate operator  $\underline{Q}$ , which, as already mentioned before, is a non-linear—superadditive—generalisation of an intensity matrix. We provide a definition, prove a number of properties, and establish a connection with lower transition operators.

Having introduced all of these related concepts and their properties, the rest of this paper focusses on our main object of interest, which is the time-dependent lower transition operator  $\underline{T}_t$  that corresponds to a given lower transition rate operator  $\underline{Q}$ . Section 6 defines this operator as the unique solution to Equation (1), shows that it is indeed a lower transition operator, and then proves that it also satisfies another—closely related—differential equation, which applies directly to  $\underline{T}_t$  rather than  $\underline{T}_t f$ . We end this section by establishing a limit expression for  $\underline{T}_t$ , which resembles—and generalises—the well-known limit expression of a matrix exponential.

With these characterisations of  $\underline{T}_t$  in hand, Section 7 then moves on to study its limit behaviour, and in particular, its ergodicity. First of all, we show that  $\underline{Q}$  is ergodic— $\lim_{t \rightarrow +\infty} \underline{T}_t f$  exists and is constant—if and only if, for any  $t > 0$ ,  $\underline{T}_t$  is ergodic in the discrete-time sense of Section 4. Secondly, for any  $t > 0$ , we show that  $\underline{T}_t$  is

regularly absorbing if and only if it is 1-step absorbing. Thirdly, we establish a simple qualitative method for checking whether  $\underline{T}_t$  is 1-step absorbing; this method does not depend on  $t$ , and is expressed directly in terms of the lower transition rate operator  $\underline{Q}$ . Finally, we explain how these three results, when combined, lead to a simple necessary and sufficient condition for  $\underline{Q}$  to be ergodic. All that is needed in order to check this condition, is the sign of a limited number of evaluations of  $\underline{Q}$ .

Section 8 concludes this paper. It briefly discusses our main result and then goes on to suggest some ideas for future research, including a number of specific conjectures and open questions that we consider to be important. The proofs of all our results are gathered in Appendix A; they are organised per section and in order of appearance. The appendix also contains some additional technical lemmas.

## 2 Preliminaries

Consider some finite *state space*  $\mathcal{X}$ . Let  $\mathcal{L}(\mathcal{X})$  be the set of all real-valued functions on  $\mathcal{X}$ . For any  $S \in \mathcal{X}$ , let  $\mathbb{I}_S \in \mathcal{L}(\mathcal{X})$  be the indicator of  $S$ , defined by  $\mathbb{I}_S(x) := 1$  if  $x \in S$  and  $\mathbb{I}_S(x) := 0$  otherwise. If  $S$  is a singleton  $\{x\}$ , we also write  $\mathbb{I}_x$  instead of  $\mathbb{I}_{\{x\}}$ . We use  $I$  to denote the identity map that maps any  $f \in \mathcal{L}(\mathcal{X})$  to itself.  $\mathbb{N}$  is the set of natural numbers without zero and  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ .

For any  $f \in \mathcal{L}(\mathcal{X})$ , we let  $\|f\| := \|f\|_\infty := \max\{|f(x)| : x \in \mathcal{X}\}$  be the maximum norm. For any operator  $A$  from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$  that is non-negatively homogeneous, meaning that

$$A(\lambda f) = \lambda A(f) \text{ for all } f \in \mathcal{L}(\mathcal{X}) \text{ and all } \lambda \geq 0,$$

we consider the induced operator norm

$$\|A\| := \sup\{\|Af\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\}. \quad (2)$$

Not only do these norms satisfy the usual defining properties of a norm, they also satisfy the following additional properties; see Appendix A.1 for a proof. For all  $f \in \mathcal{L}(\mathcal{X})$  and all operators  $A, B$  from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$  that are non-negatively homogeneous:

$$\text{N1: } \|Af\| \leq \|A\| \|f\|$$

$$\text{N2: } \|AB\| \leq \|A\| \|B\|$$

## 3 Lower transition operators

The first type of non-negatively homogeneous operator that we will consider in this paper is a lower transition operator  $\underline{T}$ . As we will show in Section 6, the solution  $\underline{T}_t$  of the differential equation that we study in this paper is of this type.

**Definition 1** (Lower transition operator). *A lower transition operator  $\underline{T}$  is a map from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$  such that for all  $f, g \in \mathcal{L}(\mathcal{X})$  and  $\lambda \geq 0$ :*

$$\text{L1: } \underline{T}f \geq \min f;$$

$$\text{L2: } \underline{T}(f + g) \geq \underline{T}(f) + \underline{T}(g); \quad [\text{superadditivity}]$$

$$\text{L3: } \underline{T}(\lambda f) = \lambda \underline{T}(f). \quad [\text{non-negative homogeneity}]$$

The corresponding upper transition operator  $\overline{T}$  is defined by

$$\overline{T}f := -\underline{T}(-f) \text{ for all } f \in \mathcal{L}(\mathcal{X}). \quad (3)$$

Basically, a lower transition operator is just a superadditive generalisation of a stochastic matrix. If the superadditivity axiom is replaced by an additivity axiom, a lower transition operator will coincide with its upper transition operator, and can then be identified with a stochastic matrix  $T$ .

For every lower transition operator  $\underline{T}$  and any  $x \in \mathcal{X}$ , the operator  $\underline{T}(\cdot)(x)$  is a coherent lower prevision [10, 11]: a superadditive, non-negatively homogeneous map from  $\mathcal{L}(\mathcal{X})$  to  $\mathbb{R}$  that dominates the min-operator. Therefore, lower transition operators are basically just finite vectors of coherent lower previsions. As a direct consequence, the following properties are implied by the corresponding versions for coherent lower previsions; see Reference [11, 2.6.1]. For any  $f, g \in \mathcal{L}(\mathcal{X})$  and  $\mu \in \mathbb{R}$  and all sequences  $\{f_n\}_{n \in \mathbb{N}} \subseteq \mathcal{L}(\mathcal{X})$ :

- L4:  $\min f \leq \underline{T}f \leq \overline{T}f \leq \max f$ ;
- L5:  $\underline{T}(f + \mu) = \underline{T}(f) + \mu$ ;
- L6:  $f \geq g \Rightarrow \underline{T}(f) \geq \underline{T}(g)$  and  $\overline{T}(f) \geq \overline{T}(g)$ ;
- L7:  $|\underline{T}f - \underline{T}g| \leq \overline{T}(|f - g|)$ ;
- L8:  $f_n \rightarrow f \Rightarrow \underline{T}f_n \rightarrow \underline{T}f$ .

As a rather straightforward consequence of L4 and L7, we also find that

- L9:  $\|\underline{T}\| \leq 1$ ;
- L10:  $\|\underline{T}f - \underline{T}g\| \leq \|f - g\|$ ;
- L11:  $\|\underline{T}A - \underline{T}B\| \leq \|A - B\|$ ,

where  $A$  and  $B$  are non-negatively homogeneous operators from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$ ; see Appendix A.2 for a proof. Finally, as this next result establishes, a sequence of lower transition operators converges pointwise if and only if it converges with respect to the operator norm.

**Proposition 1.** *For any lower transition operator  $\underline{T}$  and any sequence  $\{\underline{T}_n\}_{n \in \mathbb{N}}$  of lower transition operators:*

$$\underline{T}_n \rightarrow \underline{T} \Leftrightarrow \forall f \in \mathcal{L}(\mathcal{X}): \underline{T}_n f \rightarrow \underline{T}f.$$

## 4 Ergodicity for lower transition operators

In the linear case, that is, if the lower transition operator  $\underline{T}$  is actually a stochastic matrix  $T$ , then under rather weak assumptions,  $T^n$  converges to a limit matrix that has identical rows, or equivalently, for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\lim_{n \rightarrow +\infty} T^n f$  exists and is a constant function. This property of  $T$  is called ergodicity,<sup>3</sup> and the conditions under which it happens are well-studied; see for example Reference [13, Section 4.2].

<sup>3</sup> This terminology is not universally adopted; we follow Seneta [13, p.128]. Some authors use ergodicity to refer to a stronger property, which additionally requires that the identical rows of  $\lim_{n \rightarrow +\infty} T^n$  consist of strictly positive elements, and which can be shown to be equivalent to the existence of some  $n \in \mathbb{N}$  such that  $T^n$  consists of strictly positive elements only.

For our present purposes, we are interested in a generalised version of this concept of ergodicity, which applies to lower transition operators.

**Definition 2** (Ergodic lower transition operator). *A lower transition operator  $\underline{T}$  is ergodic if, for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\lim_{n \rightarrow \infty} \underline{T}^n f$  exists and is a constant function.*

Similarly, the corresponding upper transition operator  $\overline{T}$  is said to be ergodic if, for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\lim_{n \rightarrow \infty} \overline{T}^n f$  exists and is a constant function. It follows from Equation (3) that both notions are equivalent:  $\underline{T}$  is ergodic if and only if  $\overline{T}$  is.

Hermans and De Cooman characterised this notion of ergodicity in Reference [12], showing that a lower transition operator is ergodic if and only if it is regularly absorbing; see Proposition 2 further on. The following definition of a regularly absorbing lower transition operator is an equivalent but slightly simplified version of theirs; Lemma 16 in Appendix A.3 establishes the equivalence.

**Definition 3** (Regularly absorbing lower transition operator). *A lower transition operator  $\underline{T}$  is regularly absorbing if it satisfies the following two conditions:*

$$\mathcal{X}_{\text{RA}} := \{x \in \mathcal{X} : (\exists n \in \mathbb{N}) \min \overline{T}^n \mathbb{I}_x > 0\} \neq \emptyset$$

and

$$(\forall x \in \mathcal{X} \setminus \mathcal{X}_{\text{RA}}) (\exists n \in \mathbb{N}) \underline{T}^n \mathbb{I}_{\mathcal{X}_{\text{RA}}}(x) > 0.$$

*The first condition is called top class regularity and the second condition is called top class absorption.*

**Proposition 2.** *A lower transition operator  $\underline{T}$  is ergodic if and only if it is regularly absorbing.*

If a lower transition operator satisfies Definition 3 with  $n := 1$ , we call this lower transition operator 1-step absorbing.

**Definition 4** (1-step absorbing lower transition operator). *A lower transition operator  $\underline{T}$  is 1-step absorbing if it satisfies the following two conditions:*

$$\mathcal{X}_{1\text{A}} := \{x \in \mathcal{X} : \min \overline{T} \mathbb{I}_x > 0\} \neq \emptyset$$

and

$$(\forall x \in \mathcal{X} \setminus \mathcal{X}_{1\text{A}}) \underline{T} \mathbb{I}_{\mathcal{X}_{1\text{A}}}(x) > 0.$$

Since  $\mathcal{X}_{1\text{A}}$  is clearly subset of  $\mathcal{X}_{\text{RA}}$ , it follows from L6 that  $\underline{T} \mathbb{I}_{\mathcal{X}_{\text{RA}}} \geq \underline{T} \mathbb{I}_{\mathcal{X}_{1\text{A}}}$ , and therefore, every 1-step absorbing lower transition operator is guaranteed to be regularly absorbing as well. By combining this observation with Proposition 2, it follows that being 1-step absorbing is a sufficient condition for ergodicity. However, in general, this stronger condition of being 1-step absorbing is not necessary for ergodicity. The reason why we are nevertheless interested in this stronger property is because, as we will show further on in Section 7, for the particular lower transition operators  $\underline{T}_i$  that are the focus of this paper, both of these properties—Definitions 3 and 4—are equivalent; see Proposition 11.

## 5 Lower transition rate operators

Having introduced a non-linear generalisation of a stochastic matrix, we now move on to introduce a similar generalisation of an intensity matrix—a matrix that has non-negative off-diagonal elements and rows that sum to zero. Again, the only difference is the additivity axiom, which we relax by replacing it with a superadditivity axiom.

**Definition 5** (Lower transition rate operator). *A lower transition rate operator  $\underline{Q}$  is a map from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$  such that for all  $f, g \in \mathcal{L}(\mathcal{X})$ ,  $\lambda \geq 0$ ,  $\mu \in \mathbb{R}$  and  $x, y \in \mathcal{X}$ :*

- R1:  $\underline{Q}(\mu) = 0$ ;
- R2:  $\underline{Q}(f + g) \geq \underline{Q}(f) + \underline{Q}(g)$ ; [superadditivity]
- R3:  $\underline{Q}(\lambda f) = \lambda \underline{Q}(f)$ ; [non-negative homogeneity]
- R4:  $x \neq y \Rightarrow \underline{Q}(\mathbb{I}_y)(x) \geq 0$ .

The corresponding upper transition operator  $\overline{Q}$  is defined by

$$\overline{Q}f := -\underline{Q}(-f) \text{ for all } f \in \mathcal{L}(\mathcal{X}). \quad (4)$$

As a rather straightforward consequence of this definition, a lower transition rate operator also satisfies the following properties; see Appendix A.4 for a proof. For all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\mu \in \mathbb{R}$  and  $x \in \mathcal{X}$ :

- R5:  $\underline{Q}(f) \leq \overline{Q}(f)$ ;
- R6:  $\underline{Q}(f + \mu) = \underline{Q}(f)$ ;
- R7:  $\overline{Q}(\mathbb{I}_x)(x) \leq 0$ ;
- R8:  $2\|f\| \underline{Q}(\mathbb{I}_x)(x) \leq (f(x) - \min f) \underline{Q}(\mathbb{I}_x)(x) \leq \underline{Q}(f)(x)$ ;
- R9:  $\|\underline{Q}\| \leq 2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)|$ .

Lower transition rate operators are very closely related to lower transition operators: they can be derived from each other. The following two results make this explicit.

**Proposition 3.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for all  $\Delta \geq 0$  such that  $\Delta \|\underline{Q}\| \leq 1$ ,  $I + \Delta \underline{Q}$  is a lower transition operator.*

**Proposition 4.** *Let  $\underline{T}$  be a lower transition operator. Then for all  $\Delta > 0$ ,  $\underline{Q} := 1/\Delta(\underline{T} - I)$  is a lower transition rate operator.*

Because of this connection, we can use results for lower transition operators to obtain similar results for lower transition rate operators. The following properties can for example be derived from L8, L10 and L11 respectively; see Appendix A.4 for a proof. For any sequence  $\{f_n\}_{n \in \mathbb{N}} \subseteq \mathcal{L}(\mathcal{X})$  and all  $f, g \in \mathcal{L}(\mathcal{X})$ :

- R10:  $f_n \rightarrow f \Rightarrow \underline{Q}f_n \rightarrow \underline{Q}f$ ;
- R11:  $\|\underline{Q}f - \underline{Q}g\| \leq 2\|\underline{Q}\| \|f - g\|$ ;
- R12:  $\|\underline{Q}A - \underline{Q}B\| \leq 2\|\underline{Q}\| \|A - B\|$ ,

where  $A$  and  $B$  are non-negatively homogeneous operators from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$ . Similarly, the following result can be derived from Proposition 1.

**Proposition 5.** *For any lower transition rate operator  $\underline{Q}$  and any sequence  $\{\underline{Q}_n\}_{n \in \mathbb{N}}$  of lower transition rate operators:*

$$\underline{Q}_n \rightarrow \underline{Q} \Leftrightarrow \forall f \in \mathcal{L}(\mathcal{X}): \underline{Q}_n f \rightarrow \underline{Q} f.$$

## 6 The differential equation of interest

With all of the above material in place, we are now ready to introduce our main object of study: the time-dependent operator  $\underline{T}_t$  that corresponds to a given lower transition rate operator.

Let  $\underline{Q}$  be an arbitrary lower transition rate operator. Then for any  $t \geq 0$ , we let  $\underline{T}_t$  be a map from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$ , defined for all  $f \in \mathcal{L}(\mathcal{X})$  by the differential equation

$$\frac{d}{dt} \underline{T}_t f = \underline{Q} \underline{T}_t f \quad \text{for all } t \geq 0 \quad (5)$$

and the boundary condition  $\underline{T}_0 f := f$ . This definition is justified by a recent result of Škulj [3], who showed that the above differential equation has a unique solution for all  $t \geq 0$ .

If  $\underline{Q}$  is additive, or equivalently, if  $\underline{Q}$  can be identified with an intensity matrix  $Q$ , then  $\underline{T}_t$  is equal to its matrix exponential  $e^{Qt}$ . In the general case, the operator  $\underline{T}_t$  can be regarded as a superadditive generalisation of the matrix exponential. The rest of this section presents a number of basic properties of this operator and establishes some alternative characterisations for it.

First of all, as a direct consequence of its definition, we find that  $\underline{T}_t$  satisfies the following semigroup property:

$$\underline{T}_{t_1+t_2} = \underline{T}_{t_1} \underline{T}_{t_2} \quad \text{for all } t_1, t_2 \geq 0. \quad (6)$$

Secondly, as already suggested by our notation,  $\underline{T}_t$  is a lower transition operator.

**Proposition 6.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for all  $t \geq 0$ ,  $\underline{T}_t$  is a lower transition operator.*

Thirdly, as our next result establishes, we do not need to consider the above differential equation for every  $f \in \mathcal{L}(\mathcal{X})$  separately. Instead, we can apply a similar differential equation to the operator  $\underline{T}_t$  itself.

**Proposition 7.** *Let  $\underline{Q}$  be a lower transition rate operator. Then  $\underline{T}_0 = I$  and*

$$\frac{d}{dt} \underline{T}_t = \underline{Q} \underline{T}_t \quad \text{for all } t \geq 0, \quad (7)$$

where the derivative is taken with respect to the operator norm.

Finally,  $\underline{T}_t$  can also be defined directly, without any reference to a differential equation. The following simple limit expression resembles—and generalises—the well-known limit definition of a matrix exponential.



**Proposition 8.** *Let  $\underline{Q}$  be a lower transition rate operator. Then*

$$\underline{T}_t = \lim_{n \rightarrow \infty} \left( I + \frac{t}{n} \underline{Q} \right)^n$$

for all  $t \geq 0$ .

The operator  $\underline{T}_t$  also satisfies some additional properties, some of which are stated and proved in Appendices A.5 and A.6. However, since these properties are rather technical, and because we only need them in our proofs, we have chosen not to include them in the main text. Nevertheless, some of these properties—especially those that are stated in Proposition 25 and Corollary 26—may be of independent interest to the reader.

## 7 Ergodicity for lower transition rate operators

Having introduced our main object of study in the previous section, we now move on to study its limit behaviour and, in particular, the conditions under which  $\underline{Q}$  is ergodic. In the linear case, that is, if  $\underline{Q}$  can be identified with an intensity matrix  $\underline{Q}$ , then  $\underline{Q}$  is said to be ergodic if  $e^{\underline{Q}t}$  converges to a matrix that has identical rows,<sup>4</sup> or equivalently, if for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\lim_{t \rightarrow +\infty} e^{\underline{Q}t} f$  exists and is a constant function. We generalise this property to the non-linear case in the following way.

**Definition 6** (Ergodic lower transition rate operator). *A lower transition rate operator  $\underline{Q}$  is ergodic if, for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\lim_{t \rightarrow \infty} \underline{T}_t f$  exists and is a constant function.*

As we explained in the introduction, this property is particularly important in the context of imprecise continuous-time Markov chains, as it can then be interpreted to mean that such an imprecise continuous-time Markov chain converges to a unique limiting distribution that does not depend on the initial state. However, for the purposes of this paper, it is not necessary to understand the details of this interpretation. Instead, we will regard ergodicity as a purely mathematical property. The main contribution of this section—and, more generally, this paper—is a simple necessary and sufficient condition for a lower transition operator  $\underline{Q}$  to be ergodic.

Our first step towards finding this condition is to link the continuous-time type of ergodicity that is considered in Definition 6 to the discrete-time version that we discussed in Section 4. Our next result establishes that  $\underline{Q}$  is ergodic in the sense of Definition 6 if and only if, for some arbitrary but fixed time  $t > 0$ , the operator  $\underline{T}_t$  is ergodic in the sense of Definition 2.

**Proposition 9.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t > 0$ ,  $\underline{Q}$  is ergodic if and only if  $\underline{T}_t$  is ergodic.*

At first sight—at least to us—this result is rather surprising. Since the ergodicity of  $\underline{Q}$  is a property that depends on the evolution of  $\underline{T}_t$  as  $t$  approaches infinity, one would not suspect such a property to be completely determined by the features of a single operator  $\underline{T}_t$ , on an arbitrary time point  $t > 0$ . Nevertheless, as the above result shows, this is indeed the case.

By combining this result with Proposition 2, we immediately obtain the following alternative characterisation of ergodicity.

<sup>4</sup> Again, as was the case for the discrete-time version that we discussed in Section 4, our use of this terminology is not universally adopted; our definition is equivalent to that of Tornambè [14, Definition 4.17]. There are also authors who use ergodicity to refer to a stronger property, which additionally requires that the identical rows of  $\lim_{t \rightarrow +\infty} e^{\underline{Q}t}$  consist of strictly positive elements.

**Corollary 10.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t > 0$ ,  $\underline{Q}$  is ergodic if and only if  $\underline{T}_t$  is regularly absorbing.*

This result is clearly a good first step in obtaining a simple characterisation of ergodicity. Indeed, due to this result, instead of having to compute—or approximate—the limit behaviour of  $\underline{T}_t$  as  $t$  approaches infinity, it now suffices to restrict attention to a single time point  $t > 0$ , which we can even choose ourselves, and to check whether for this time point  $t$ , the operator  $\underline{T}_t$  is regularly absorbing. Furthermore, as our next result establishes, checking whether this particular type of lower transition operator is regularly absorbing is easier than it is for general lower transition operators: in this special case, being regularly absorbing is equivalent to being 1-step absorbing.

**Proposition 11.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t \geq 0$ ,  $\underline{T}_t$  is regularly absorbing if and only if it is 1-step absorbing.*

By combining this result with Corollary 10, we immediately obtain yet another necessary and sufficient condition for  $\underline{Q}$  to be ergodic.

**Corollary 12.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t > 0$ ,  $\underline{Q}$  is ergodic if and only if  $\underline{T}_t$  is 1-step absorbing.*

Because of this result, checking whether  $\underline{Q}$  is ergodic is now reduced to checking whether  $\underline{T}_t$  is 1-step absorbing, for some arbitrary but fixed  $t > 0$ . Although this is already easier than studying the limit behaviour of  $\underline{T}_t$  directly, it is still non-trivial. As can be seen from Definition 4, it requires us to evaluate the strict positivity of numbers that are of the form  $\overline{T}_t \mathbb{I}_x(y)$  and  $\underline{T}_t \mathbb{I}_A(x)$ , with  $x, y \in \mathcal{X}$  and  $A \subseteq \mathcal{X}$ . At first sight, this still seems to be a rather cumbersome task that will involve either solving the differential Equation (5) or applying the limit expression in Proposition 8. However, as it turns out, this is not the case.

Indeed, as we are about to show, the strict positivity of  $\overline{T}_t \mathbb{I}_x(y)$  and  $\underline{T}_t \mathbb{I}_A(x)$  does not depend on the specific value of  $t$ , but only on the lower transition operator  $\underline{Q}$ . In order to make this specific, we introduce the following notions of upper and lower reachability.

**Definition 7** (Upper reachability). *For any  $x, y \in \mathcal{X}$ , we say that  $x$  is upper reachable from  $y$ , and denote this by  $y \rightarrow x$ , if there is some sequence  $y = x_0, \dots, x_n = x$  such that, for all  $k \in \{1, \dots, n\}$ :*

$$x_k \neq x_{k-1} \text{ and } \overline{Q}(\mathbb{I}_{x_k})(x_{k-1}) > 0.$$

**Definition 8** (Lower reachability). *For any  $x \in \mathcal{X}$  and  $A \subseteq \mathcal{X}$ , we say that  $A$  is lower reachable from  $x$ , and denote this by  $x \rightarrow A$ , if  $x \in A_n$ , where  $\{A_k\}_{k \in \mathbb{N}_0}$  is an increasing sequence that is defined by  $A_0 := A$  and*

$$A_{k+1} := A_k \cup \{y \in \mathcal{X} \setminus A_k : \underline{Q}(\mathbb{I}_{A_k})(y) > 0\} \text{ for all } k \in \mathbb{N}_0, \quad (8)$$

*and where  $n$  is the first index such that  $A_n = A_{n+1}$ .*

An important property of both of these two notions is that they are easy to check. For upper reachability, it suffices to draw a directed graph that has the elements of  $\mathcal{X}$  as its nodes and which features an arrow from  $y$  to  $x$  if and only if  $\overline{Q}(\mathbb{I}_x)(y) > 0$ . Checking whether  $y$  is upper reachable from  $x$  is then clearly equivalent to checking whether it is possible to start in  $x$  and follow the arrows in the graph to reach  $y$ . This is a standard reachability problem that can either be solved manually, or by means of techniques

from graph theory. Lower reachability essentially requires us to construct the sequence  $\{A_k\}_{k \in \mathbb{N}_0}$  up to the index  $n$ . Since it follows from the increasing nature of this sequence that  $n \leq |\mathcal{X} \setminus A|$ , this too is a straightforward task.

The reason why we are interested in these notions of lower and upper reachability are the following two equivalences.

**Proposition 13.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t > 0$  and any  $x, y \in \mathcal{X}$ :*

$$\overline{T}_t \mathbb{I}_x(y) > 0 \Leftrightarrow y \dot{\rightarrow} x.$$

**Proposition 14.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for any  $t > 0$ , any  $x \in \mathcal{X}$  and any  $A \subseteq \mathcal{X}$ :*

$$\underline{T}_t \mathbb{I}_A(x) > 0 \Leftrightarrow x \dot{\rightarrow} A.$$

By combining these equivalences with Definition 4 and Corollary 12, we easily obtain the following result, which is the characterisation of ergodicity that we have been after all along.

**Theorem 15.** *A lower transition rate operator  $\underline{Q}$  is ergodic if and only if*

$$\mathcal{X}_{1A} := \{x \in \mathcal{X} : (\forall y \in \mathcal{X}) y \dot{\rightarrow} x\} \neq \emptyset$$

and

$$(\forall x \in \mathcal{X} \setminus \mathcal{X}_{1A}) x \dot{\rightarrow} \mathcal{X}_{1A}.$$

We consider this necessary and sufficient condition for the ergodicity of  $\underline{Q}$  to be the main contribution of this paper. The reason why it is to be preferred over other necessary and sufficient conditions, such as those that are given in Corollaries 10 and 12, is because it does not require us to evaluate the operator  $\underline{T}_t$ . Instead, all we have to do is solve a limited number of lower and upper reachability problems, which, as can be seen from Definitions 7 and 8, only requires us to evaluate the operator  $\underline{Q}$ . This is obviously preferable, because  $\underline{Q}$  is directly available, whereas  $\underline{T}_t$  is known only indirectly through the differential Equation (5) or the limit expression in Proposition 8.

## 8 Conclusions

The main contribution of this paper is a simple necessary and sufficient condition for the ergodicity of a lower transition rate operator  $\underline{Q}$ . Specifically, as can be seen from Theorem 15, it is necessary and sufficient for at least one state  $x$  to be upper reachable from every other state  $y$ , and for the set  $\mathcal{X}_{1A}$  of all the states that satisfy this condition to be lower reachable from each of the states that does not. The main conclusion that can be drawn from this result is that ergodicity is easily satisfied. For example, it already suffices—but is by no means necessary—for every state to be upper reachable from any other state.

This result provides us with a simple method for checking whether  $\underline{Q}$  is ergodic, or equivalently, whether  $\underline{T}_t f$  is guaranteed to converge to a constant function as  $t$  approaches infinity. The reason why this is important to us is because, as explained in the introduction, in the context of imprecise continuous-time Markov chains, this notion of ergodicity is equivalent to the existence of a unique limiting distribution that is independent of the initial state. Therefore, our results can be used to check whether or not such a unique limiting distribution exists.

Although the existence of such a limiting distribution is important, it is of course only one of the many aspects of the limit behaviour of imprecise continuous-time Markov chains. Many problems still remain unsolved. For example: what happens if we no longer care about the influence of the initial state? In particular: for a given initial state, under which conditions will an imprecise continuous-time Markov chain converge to a limiting distribution that is allowed to depend on this initial state? Or equivalently, using the terminology of this paper: which conditions does  $\underline{Q}$  need to satisfy in order for  $\lim_{t \rightarrow +\infty} \underline{T}_t f(x)$  to exist? Ergodicity is clearly a sufficient condition—since it additionally requires that this limit does not depend of  $x$ —but it is definitely not necessary. In fact, we even conjecture that this type of convergence requires no conditions at all.

The simple reason why we suspect this conjecture to hold is because, as mentioned in the introduction, if  $\underline{Q}$  is an intensity matrix  $Q$ , then rather remarkably, regardless of the specific intensity matrix  $Q$  that is considered,  $\underline{T}_t = e^{Q_t}$  will always converge to a limit [1, Theorem II.10.1]. By analogy, for any lower transition rate operator  $\underline{Q}$ , we think that the corresponding lower transition operator  $\underline{T}_t$  will always converge to a limit. Investigating whether this conjecture is indeed true would be a nice topic for future research.

Another interesting line of future research would be to study ergodicity—or other types of convergence—from a quantitative rather than just qualitative point of view, by also taking into account the rate of convergence. For the discrete-time type of ergodicity that we discussed in Section 4, such a study has already been conducted in References [12, 15], leading to the development of a coefficient of ergodicity that simultaneously captures both the qualitative aspect of convergence—“does it converge or not?”—and the quantitative aspect—“at which rate does it converge?”. We think that similar coefficients of ergodicity can also be developed for the continuous-time models that we have considered in this paper.

Finally, we would like to point out that these suggestions for future research are just the tip of the iceberg, because they focus solely on the limit behaviour of imprecise continuous-time Markov chains. Ultimately, we hope that our contributions will serve as a first step towards a further theoretic development of the general field of imprecise continuous-time Markov chains. The reason why we consider such developments to be important is because, given the success of precise continuous-time Markov chains in various fields of application [2], and the ever increasing demand for features such as reliability and robustness in these applications, we are convinced that imprecise continuous-time Markov chains have plenty of applied potential. Nevertheless, almost no applications have been developed so far. It seems to us that one of the main reasons for this lack of applications is a severe lack of available theoretical tools. We hope that a further theoretical development of the field of imprecise continuous-time Markov chains will allow this field to flourish, and will turn it into a full-fledged robust extension of the field of continuous-time Markov chains.

## Acknowledgements

Jasper De Bock is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO) and wishes to acknowledge its financial support. The author would also like to thank Gert de Cooman, Matthias C. M. Troffaes, Stavros Lopotatzidis and Thomas Krak for stimulating discussions on the topic of imprecise continuous-time Markov chains.

## References

- [1] K. L. Chung, Markov chains with stationary transition probabilities, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer, Berlin, New York, 1967.
- [2] W. J. Anderson, Continuous-Time Markov chains, an Applications-Oriented Approach, Springer Series in Statistics, Springer, New York, 1991.
- [3] D. Škulj, Efficient computation of the bounds of continuous time imprecise Markov chains, Applied mathematics and computation 250 (C) (2015) 165–180.
- [4] T. Krak, J. De Bock, Imprecise Continuous-Time Markov Chains Work in progress.
- [5] M. C. M. Troffaes, J. Gledhill, D. Škulj, S. Blake, Using imprecise continuous time markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair, in: ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications, 2015, pp. 287–294.
- [6] D. Škulj, Interval matrix differential equations (2012). [arXiv:1204.0467](https://arxiv.org/abs/1204.0467).
- [7] X. Guo, O. Hernández-Lerma, Continuous-time Markov decision processes: theory and applications., Stochastic Modelling and Applied Probability 62, Springer, Berlin, 2009.
- [8] X. Guo, O. Hernndez-Lerma, Continuous-time controlled markov chains, Annals of Applied Probability 13 (1) (2003) 363–388.
- [9] S. Galdino, Interval continuous-time markov chains simulation, in: Proceedings of the 1013 International Conference on Fuzzy Theory and Its Applications, 2013, pp. 273–278.
- [10] M. C. M. Troffaes, G. de Cooman, Lower previsions, John Wiley & Sons, 2014.
- [11] P. Walley, Statistical reasoning with imprecise probabilities, Chapman and Hall, London, 1991.
- [12] F. Hermans, G. de Cooman, Characterisation of ergodic upper transition operators, International Journal of Approximate Reasoning 53 (4) (2012) 573–583.
- [13] E. Seneta, Non-negative matrices and Markov chains, Springer, New York, 2006.
- [14] A. Tornambè, Discrete-event System Theory: An Introduction, World Scientific, 1995.
- [15] D. Škulj, R. Hable, Coefficients of ergodicity for Markov chains with uncertain parameters, Metrika 76 (1) (2011) 107–133.
- [16] H. L. Royden, P. M. Fitzpatrick, Real Analysis, 4th Edition, Prentice Hall, 2010.

## A Proofs

### A.1 Proofs of results in Section 2

Let  $A$  and  $B$  be two non-negatively homogeneous operators from  $\mathcal{L}(\mathcal{X})$  to  $\mathcal{L}(\mathcal{X})$  and consider any  $f, g \in \mathcal{L}(\mathcal{X})$  and  $\lambda \in \mathbb{R}$ .

It is well known that the maximum norm on  $\mathcal{L}(\mathcal{X})$  satisfies the defining properties of a norm: it is absolutely homogeneous ( $\|\lambda f\| = |\lambda| \|f\|$ ), it is subadditive ( $\|f + g\| \leq \|f\| + \|g\|$ ) and it separates points ( $\|f\| = 0 \Rightarrow f = 0$ ). The induced operator norm also satisfies these properties. Firstly, it is absolutely homogeneous because the maximum norm is:

$$\begin{aligned} \|\lambda A\| &= \sup\{\|\lambda Af\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &= \sup\{|\lambda| \|Af\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &= |\lambda| \sup\{\|Af\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} = |\lambda| \|A\|. \end{aligned}$$

Secondly, it is subadditive because the maximum norm is:

$$\begin{aligned} \|A + B\| &= \sup\{\|(A + B)f\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &= \sup\{\|Af + Bf\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &\leq \sup\{\|Af\| + \|Bf\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &\leq \sup\{\|A\| + \|B\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} = \|A\| + \|B\|. \end{aligned}$$

Thirdly, it separates points because the maximum norm does: if  $\|A\| = 0$ , then  $A = 0$  because, for all  $f \in \mathcal{L}(\mathcal{X})$ , it follows from N1—which we will prove next—that

$$0 \leq \|Af\| \leq \|A\| \|f\| = 0$$

and therefore, since the maximum norm separates points, that  $Af = 0$ .

In order to prove N1, we consider two cases:  $f = 0$  and  $f \neq 0$ . If  $f \neq 0$ , or equivalently, if  $\|f\| \neq 0$ , we let  $g := f/\|f\|$ . If  $f = 0$ , or equivalently, if  $\|f\| = 0$ , we let  $g := 1$ . In both cases, this guarantees that  $f = \|f\| g$  and  $\|g\| = 1$  and therefore, we find that

$$\|Af\| = \|A(\|f\| g)\| = \|\|f\| Ag\| = \|f\| \|Ag\| \leq \|f\| \|A\|,$$

where the inequality holds because  $\|f\| \geq 0$  and  $\|Ag\| \leq \|A\|$ .

Finally, N2 follows rather easily from N1:

$$\begin{aligned} \|AB\| &= \sup\{\|ABf\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &\leq \sup\{\|A\| \|Bf\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} \\ &= \|A\| \sup\{\|Bf\| : f \in \mathcal{L}(\mathcal{X}), \|f\| = 1\} = \|A\| \|B\|. \end{aligned}$$

### A.2 Proofs of results in Section 3

*Proof of L9, L10 and L11.* L9 follows from Equation (2) because we know from L4 that  $\|\underline{T}f\| \leq \|f\|$  for all  $f \in \mathcal{L}(\mathcal{X})$ . L10 follows from L7 and L4 (in that order). L11 follows from Equation (2) and L10.  $\square$

*Proof of Proposition 1.* The direct implication follows trivially from N1. For the converse implication, we provide a proof by contradiction. Assume that  $\underline{T}_n f \rightarrow \underline{T}f$  for all  $f \in \mathcal{L}(\mathcal{X})$ . Assume *ex absurdo* that  $\underline{T}_n \not\rightarrow \underline{T}$ . Then since  $\underline{T}_n \not\rightarrow \underline{T}$ , it follows that

$\limsup_{n \rightarrow \infty} \|\underline{T}_n - \underline{T}\| > 0$ , which implies that there is some  $\varepsilon > 0$  and an increasing sequence  $n_k, k \in \mathbb{N}$ , of natural numbers such that  $\|\underline{T}_{n_k} - \underline{T}\| > \varepsilon$  for all  $k \in \mathbb{N}$ . Furthermore, for all  $k \in \mathbb{N}$ , it follows from  $\|\underline{T}_{n_k} - \underline{T}\| > \varepsilon$  and Equation (2) that there is some  $f_k \in \mathcal{L}(\mathcal{X})$  such that  $\|f_k\| = 1$  and  $\|\underline{T}_{n_k} f_k - \underline{T} f_k\| > \varepsilon$ . Since the sequence  $f_k, k \in \mathbb{N}$ , is clearly bounded—because  $\|f_k\| = 1$ —it follows from the Bolzano-Weierstrass theorem that it has a convergent subsequence, which implies that there is some  $f \in \mathcal{L}(\mathcal{X})$  and an increasing sequence  $k_i, i \in \mathbb{N}$ , of natural numbers such that  $\lim_{i \rightarrow \infty} \|f_{k_i} - f\| = 0$ . Furthermore, since we have assumed that  $\underline{T}_n f \rightarrow \underline{T} f$ , it follows that

$$\lim_{i \rightarrow \infty} \|\underline{T}_{n_{k_i}} f - \underline{T} f\| = \lim_{n \rightarrow \infty} \|\underline{T}_n f - \underline{T} f\| = 0.$$

Hence, since it follows from L10 that

$$\begin{aligned} \|\underline{T}_{n_{k_i}} f_{k_i} - \underline{T} f_{k_i}\| &= \|(\underline{T}_{n_{k_i}} f_{k_i} - \underline{T}_{n_{k_i}} f) + (\underline{T}_{n_{k_i}} f - \underline{T} f) + (\underline{T} f - \underline{T} f_{k_i})\| \\ &\leq \|\underline{T}_{n_{k_i}} f_{k_i} - \underline{T}_{n_{k_i}} f\| + \|\underline{T}_{n_{k_i}} f - \underline{T} f\| + \|\underline{T} f - \underline{T} f_{k_i}\| \\ &\leq \|f_{k_i} - f\| + \|\underline{T}_{n_{k_i}} f - \underline{T} f\| + \|f_{k_i} - f\|, \end{aligned}$$

we find that

$$\lim_{i \rightarrow \infty} \|\underline{T}_{n_{k_i}} f_{k_i} - \underline{T} f_{k_i}\| = 0.$$

Since  $\|\underline{T}_{n_{k_i}} f_{k_i} - \underline{T} f_{k_i}\| > \varepsilon > 0$  for all  $i \in \mathbb{N}$ , this is a contradiction.  $\square$

### A.3 Proofs of results in Section 4

**Lemma 16.** *A lower transition operator  $\underline{T}$  is regularly absorbing if and only if*

$$\mathcal{X}'_{\text{RA}} := \{x \in \mathcal{X} : (\exists n \in \mathbb{N})(\forall k \geq n) \min \bar{T}^k \mathbb{I}_x > 0\} \neq \emptyset$$

and

$$(\forall x \in \mathcal{X} \setminus \mathcal{X}'_{\text{RA}})(\exists n \in \mathbb{N}) \bar{T}^n \mathbb{I}_{\mathcal{X} \setminus \mathcal{X}'_{\text{RA}}}(x) < 1.$$

Furthermore, the set  $\mathcal{X}'_{\text{RA}}$  is equal to the set  $\mathcal{X}_{\text{RA}}$  that was used in Definition 3.

*Proof of Lemma 16.* Consider any  $x \in \mathcal{X}_{\text{RA}}$ . Definition 3 then implies that there is some  $n \in \mathbb{N}$  such that  $\min \bar{T}^n \mathbb{I}_x > 0$ , and therefore, because of L4, we know that  $\bar{T}^{n+1} \mathbb{I}_x = \bar{T}(\bar{T}^n \mathbb{I}_x) \geq \min \bar{T}^n \mathbb{I}_x > 0$ , which implies that  $\min \bar{T}^{n+1} \mathbb{I}_x > 0$ . In the same way, we also find that  $\min \bar{T}^{n+2} \mathbb{I}_x > 0$  and, by continuing in this way, that  $\min \bar{T}^k \mathbb{I}_x > 0$  for all  $k \geq n$ . Since this holds for all  $x \in \mathcal{X}_{\text{RA}}$ , it follows that  $\mathcal{X}_{\text{RA}} \subseteq \mathcal{X}'_{\text{RA}}$ . Since  $\mathcal{X}'_{\text{RA}}$  is clearly a subset of  $\mathcal{X}_{\text{RA}}$ , this implies that  $\mathcal{X}'_{\text{RA}} = \mathcal{X}_{\text{RA}}$ . Hence, trivially,  $\mathcal{X}_{\text{RA}} \neq \emptyset$  if and only if  $\mathcal{X}'_{\text{RA}} \neq \emptyset$ . The result now follows because it holds for all  $x \in \mathcal{X} \setminus \mathcal{X}'_{\text{RA}} = \mathcal{X} \setminus \mathcal{X}_{\text{RA}}$  and all  $n \in \mathbb{N}$  that

$$\bar{T}^n \mathbb{I}_{\mathcal{X} \setminus \mathcal{X}'_{\text{RA}}}(x) = \bar{T}^n (1 - \mathbb{I}_{\mathcal{X}'_{\text{RA}}})(x) = 1 - \underline{T}^n (\mathbb{I}_{\mathcal{X}'_{\text{RA}}})(x) = 1 - \underline{T}^n (\mathbb{I}_{\mathcal{X}_{\text{RA}}})(x),$$

where the second equality follows from L5 and Equation (3).  $\square$

*Proof of Proposition 2.* Since we know from Lemma 16 that our definition of a regularly absorbing lower transition operator is equivalent to the definition in Reference [12], this result is identical to [12, Proposition 3].  $\square$

#### A.4 Proofs of results in Section 5

*Proof of R5, R6, R7, R8 and R9.* R5 holds because it follows from Equation (4), R2 and R1 that

$$\underline{Q}(f) - \overline{Q}(f) = \underline{Q}(f) + \underline{Q}(-f) \leq \underline{Q}(f - f) = \underline{Q}(0) = 0.$$

R6 holds because it follows from R1 and R2 that

$$\underline{Q}(f) = \underline{Q}(f) + \underline{Q}(\mu) \leq \underline{Q}(f + \mu) = \underline{Q}(f + \mu) + \underline{Q}(-\mu) \leq \underline{Q}(f).$$

R7 holds because it follows from Equation (4), R6, R2 and R4—in that order—that

$$\begin{aligned} \overline{Q}(\mathbb{I}_x)(x) &= -\underline{Q}(-\mathbb{I}_x)(x) = -\underline{Q}(1 - \mathbb{I}_x)(x) \\ &= -\underline{Q}(\sum_{y \in \mathcal{X} \setminus \{x\}} \mathbb{I}_y)(x) \leq -\sum_{y \in \mathcal{X} \setminus \{x\}} \underline{Q}(\mathbb{I}_y)(x) \leq 0. \end{aligned}$$

R8 holds because it follows from R6, R2, R3, R4, R7 and R5—in that order—that

$$\begin{aligned} \underline{Q}(f)(x) &= \underline{Q}(f - \min f)(x) \geq \sum_{y \in \mathcal{X}} \underline{Q}((f(y) - \min f) \mathbb{I}_y)(x) \\ &= \sum_{y \in \mathcal{X}} (f(y) - \min f) \underline{Q}(\mathbb{I}_y)(x) \\ &\geq (f(x) - \min f) \underline{Q}(\mathbb{I}_x)(x) \\ &\geq (\max f - \min f) \underline{Q}(\mathbb{I}_x)(x) \geq 2 \|f\| \underline{Q}(\mathbb{I}_x)(x). \end{aligned}$$

We end by proving R9. Consider any  $g \in \mathcal{L}(\mathcal{X})$  such that  $\|g\| = 1$ . It then follows from R8 and R7 that

$$\underline{Q}(g) \geq 2 \|g\| \min_{x \in \mathcal{X}} \underline{Q}(\mathbb{I}_x)(x) \geq -2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)|.$$

Similarly, since  $\|-g\| = \|g\| = 1$ , we also find that  $\underline{Q}(-g) \geq -2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)|$ . By combining these two inequalities with R5 and Equation (4), it follows that

$$-2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)| \leq \underline{Q}(g) \leq \overline{Q}(g) = -\underline{Q}(-g) \leq 2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)|,$$

which implies that  $\|\underline{Q}(g)\| \leq 2 \max_{x \in \mathcal{X}} |\underline{Q}(\mathbb{I}_x)(x)|$ . Since this is true for all  $g \in \mathcal{L}(\mathcal{X})$  such that  $\|g\| = 1$ , R9 now follows from Equation (2).  $\square$

*Proof of Proposition 3.* L2 and L3 follow trivially from R2 and R3. We only prove L1. Consider any  $f \in \mathcal{L}(\mathcal{X})$ . Then

$$\begin{aligned} (I + \Delta \underline{Q})f &= f + \Delta \underline{Q}f \\ &= f + \Delta \sum_{x \in \mathcal{X}} \mathbb{I}_x \underline{Q}(f)(x) \\ &\geq f + \Delta \sum_{x \in \mathcal{X}} (f(x) - \min f) \mathbb{I}_x \underline{Q}(\mathbb{I}_x)(x) \\ &\geq f - \Delta \sum_{x \in \mathcal{X}} (f(x) - \min f) \mathbb{I}_x \|\underline{Q}(\mathbb{I}_x)\| \\ &\geq f - \Delta \sum_{x \in \mathcal{X}} (f(x) - \min f) \mathbb{I}_x \|\underline{Q}\| \\ &= f - \Delta \|\underline{Q}\| (f - \min f) = (f - \min f)(1 - \Delta \|\underline{Q}\|) + \min f \geq \min f, \end{aligned}$$

where the first inequality follows from R8 and the third inequality follows from Equation (2) and the fact that  $\|\mathbb{I}_x\| = 1$ .  $\square$



*Proof of Proposition 4.* Simply check each of the defining properties: R1 holds because L4 implies that  $\underline{T}(\mu) = \mu$  for all  $\mu \in \mathbb{R}$ , R2 follows from L2, R3 follows from L3 and R4 follows from L1.  $\square$

*Proof of R10, R11 and R12.* R10, R11 and R12 are trivial if  $\underline{Q} = 0$ . Therefore, we may assume that  $\underline{Q} \neq 0$ , which implies that  $\|\underline{Q}\| > 0$ . Now let  $\underline{T} := I + 1/\|\underline{Q}\|\underline{Q}$ . It then follows from Proposition 3 that  $\underline{T}$  is a lower transition operator. We first prove R10. If  $f_n \rightarrow f$ , then  $\underline{T}f_n \rightarrow \underline{T}f$  because of L8. Since  $\underline{Q} = \|\underline{Q}\|(\underline{T} - I)$ , this implies that  $\underline{Q}f_n \rightarrow \underline{Q}f$ . R11 holds because

$$\begin{aligned} \|\underline{Q}f - \underline{Q}g\| &= \|\|\underline{Q}\|(\underline{T}f - f) - \|\underline{Q}\|(\underline{T}g - g)\| \\ &\leq \|\underline{Q}\| \|\underline{T}f - \underline{T}g\| + \|\underline{Q}\| \|f - g\| \leq 2\|\underline{Q}\| \|f - g\|, \end{aligned}$$

where the last inequality follows from L10. Similarly, R12 holds because

$$\begin{aligned} \|\underline{Q}A - \underline{Q}B\| &= \|\|\underline{Q}\|(\underline{T}A - A) - \|\underline{Q}\|(\underline{T}B - B)\| \\ &\leq \|\underline{Q}\| \|\underline{T}A - \underline{T}B\| + \|\underline{Q}\| \|A - B\| \leq 2\|\underline{Q}\| \|A - B\|, \end{aligned}$$

where the last inequality follows from L11.  $\square$

*Proof of Proposition 5.* The direct implication follows trivially from N1. We only prove the converse implication. Assume that  $\underline{Q}_n f \rightarrow \underline{Q}f$  for all  $f \in \mathcal{L}(\mathcal{X})$ . For all  $x \in \mathcal{X}$ , this implies that  $\underline{Q}_n(\mathbb{I}_x)(x) \rightarrow \underline{Q}(\mathbb{I}_x)(x)$ , which in turn implies that there is some  $c_x > 0$  such that  $|\underline{Q}(\mathbb{I}_x)(x)| < c_x$  and  $|\underline{Q}_n(\mathbb{I}_x)(x)| < c_x$  for all  $n \in \mathbb{N}$ . Let  $c := \max_{x \in \mathcal{X}} c_x$ . It then follows from R9 that  $\|\underline{Q}\| \leq 2c$  and  $\|\underline{Q}_n\| \leq 2c$  for all  $n \in \mathbb{N}$ . Choose any  $0 < \Delta \leq 1/2c$ . It then follows from Proposition 3 that  $\underline{T} := I + \Delta\underline{Q}$  and  $\underline{T}_n := I + \Delta\underline{Q}_n$ ,  $n \in \mathbb{N}$ , are lower transition operators. Furthermore, since  $\underline{Q}_n f \rightarrow \underline{Q}f$  for all  $f \in \mathcal{L}(\mathcal{X})$ , it follows that  $\underline{T}_n f \rightarrow \underline{T}f$  for all  $f \in \mathcal{L}(\mathcal{X})$ . By applying Proposition 1, we now find that  $\underline{T}_n \rightarrow \underline{T}$ , which implies that  $\underline{Q}_n \rightarrow \underline{Q}$  because

$$\|\underline{Q}_n - \underline{Q}\| = \frac{1}{\Delta} \|\Delta\underline{Q}_n - \Delta\underline{Q}\| = \frac{1}{\Delta} \|(I + \Delta\underline{Q}_n) - (I + \Delta\underline{Q})\| = \frac{1}{\Delta} \|\underline{T}_n - \underline{T}\|.$$

$\square$

## A.5 Proofs of results in Section 6

**Lemma 17.** *Let  $\underline{Q}$  be a lower transition rate operator. Then for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $\underline{T}_s f$  is continuously differentiable on  $[0, \infty)$ .*

*Proof.* It follows from Equation (5) that  $\underline{T}_s f$  is continuous on  $[0, \infty)$ . Therefore, since  $\underline{Q}$  is a continuous operator [R10],  $\underline{Q}\underline{T}_s f$  is also continuous on  $[0, \infty)$ . Because of Equation (5), this implies that  $\underline{T}_s f$  is continuously differentiable on  $[0, \infty)$ .  $\square$

**Lemma 18.** *Let  $\underline{Q}$  be a lower transition rate operator and let  $\Gamma(s)$  be a continuously differentiable map from  $[0, t]$  to  $\mathcal{L}(\mathcal{X})$  for which  $\frac{d}{ds}\Gamma(s) \geq \underline{Q}\Gamma(s)$  for all  $s \in [0, t]$ . Then  $\min \Gamma(t) \geq \min \Gamma(0)$ .*

*Proof.* Since  $\Gamma(s)$  is continuously differentiable on  $[0, t]$ , it follows that for every  $x \in \mathcal{X}$ ,  $\Gamma(s)(x)$  is also continuously differentiable on  $[0, t]$ , which implies that it is absolutely continuous on  $[0, t]$ . Hence, since a minimum of a finite number of absolutely continuous functions is again absolutely continuous, we find that  $\min \Gamma(s)$  is absolutely continuous on  $[0, t]$ , which implies—see Reference [16, Theorem 10, Section 6.5]—that

$\min \Gamma(s)$  has a derivative  $\frac{d}{ds} \min \Gamma(s)$  almost everywhere on  $(0, t)$ , that this derivative is Lebesgue integrable over  $[0, t]$ , and that

$$\min \Gamma(t) = \min \Gamma(0) + \int_0^t \left( \frac{d}{ds} \min \Gamma(s) \right) ds. \quad (9)$$

Consider now any  $t^* \in (0, t)$  for which  $\min \Gamma(s)$  has a derivative and consider any  $x \in \mathcal{X}$  for which  $\Gamma(t^*)(x) = \min \Gamma(t^*)$  [clearly, there is at least one such  $x$ ]. Since  $\Gamma(s)(x)$  is differentiable,  $\frac{d}{ds} \Gamma(s)(x)$  exists in  $t^*$ . Assume *ex absurdo* that  $\frac{d}{ds} \Gamma(s)(x)|_{s=t^*}$  is not equal to  $\frac{d}{ds} \min \Gamma(s)|_{s=t^*}$  or, equivalently, that  $\frac{d}{ds} (\Gamma(s)(x) - \min \Gamma(s))|_{s=t^*} \neq 0$ . Then, because  $\Gamma(s)(x) - \min \Gamma(s)$  is continuous [since  $\Gamma(s)(x)$  and  $\min \Gamma(s)$  are both (absolutely) continuous] and because  $t^* \in (0, t)$  and  $\Gamma(t^*)(x) - \min \Gamma(t^*) = 0$ , it follows that there is some  $t' \in (0, t)$  such that  $\Gamma(t')(x) - \min \Gamma(t') < 0$  or, equivalently, such that  $\Gamma(t')(x) < \min \Gamma(t')$ . Since this is clearly a contradiction, it follows that

$$\frac{d}{ds} \Gamma(s)(x)|_{s=t^*} = \frac{d}{ds} \min \Gamma(s)|_{s=t^*}. \quad (10)$$

We also have that

$$\frac{d}{ds} \Gamma(s)(x)|_{s=t^*} \geq \underline{Q}(\Gamma(t^*))(x) \geq (\Gamma(t^*)(x) - \min \Gamma(t^*)) \underline{Q}(\mathbb{I}_x)(x) = 0,$$

where the second inequality follows from R8 and the last equality follows because  $\Gamma(t^*)(x) = \min \Gamma(t^*)$ . By combining this result with Equation (10), we find that, for all  $t^* \in (0, t)$  for which  $\min \Gamma(s)$  has a derivative,  $\frac{d}{ds} \min \Gamma(s)|_{s=t^*} \geq 0$ . It therefore follows from Equation (9) that  $\min \Gamma(t) \geq \min \Gamma(0)$ .  $\square$

*Proof of Proposition 6.* We first prove L1. Consider any  $f \in \mathcal{L}(\mathcal{X})$ . It then follows from Lemma 17 that  $\underline{T}_s f$  is continuously differentiable on  $[0, t]$ . Therefore, and because of Equation (5), we infer from Lemma 18 that  $\min \underline{T}_t f \geq \min \underline{T}_0 f$ . Since  $\underline{T}_0 f = f$ , this implies that  $\min \underline{T}_t f \geq \min f$ , which in turn implies that  $\underline{T}_t f \geq \min f$ .

Let us now prove L2. Consider any  $f, g \in \mathcal{L}(\mathcal{X})$ . It follows from Lemma 17 that  $\underline{T}_s f$ ,  $\underline{T}_s g$  and  $\underline{T}_s(f + g)$  are continuously differentiable on  $[0, t]$ , which implies that  $\Gamma(s) := \underline{T}_s(f + g) - \underline{T}_s f - \underline{T}_s g$  is continuously differentiable on  $[0, t]$ . Furthermore, for all  $s \in [0, t]$ , it follows from Equation (5) and R2 that

$$\begin{aligned} \frac{d}{ds} \Gamma(s) &= \frac{d}{ds} \underline{T}_s(f + g) - \frac{d}{ds} \underline{T}_s f - \frac{d}{ds} \underline{T}_s g \\ &= \underline{Q} \underline{T}_s(f + g) - \underline{Q} \underline{T}_s f - \underline{Q} \underline{T}_s g \\ &= \underline{Q}(\Gamma(s) + \underline{T}_s f + \underline{T}_s g) - \underline{Q} \underline{T}_s f - \underline{Q} \underline{T}_s g \geq \underline{Q} \Gamma(s). \end{aligned}$$

Therefore, we infer from Lemma 18 that  $\min \Gamma(t) \geq \min \Gamma(0)$ . Since  $\Gamma(0) = \underline{T}_0(f + g) - \underline{T}_0 f - \underline{T}_0 g = 0$ , this implies that  $\min \Gamma(t) \geq 0$ , which in turn implies that  $\Gamma(t) \geq 0$  or, equivalently, that  $\underline{T}_t(f + g) \geq \underline{T}_t f + \underline{T}_t g$ .

We end by proving L3. Consider any  $f \in \mathcal{L}(\mathcal{X})$  and  $\lambda \geq 0$ . It then follows from Equation (5) and R3 that

$$\frac{d}{ds} (\lambda \underline{T}_s f) = \lambda \frac{d}{ds} \underline{T}_s f = \lambda \underline{Q} \underline{T}_s f = \underline{Q}(\lambda \underline{T}_s f) \text{ for all } s \geq 0.$$

Since we also have that  $\lambda \underline{T}_0 f = \lambda f = \underline{T}_0(\lambda f)$ , it follows that  $\lambda \underline{T}_s f$  satisfies the same differential equation and boundary condition as  $\underline{T}_s(\lambda f)$ . Since we know that this differential equation and boundary condition lead to a unique solution on  $[0, \infty)$ , it follows that  $\underline{T}_t(\lambda f) = \lambda \underline{T}_t(f)$ .  $\square$

**Lemma 19.** *Let  $\underline{Q}$  be a lower transition rate operator. Then*

$$\lim_{\Delta \rightarrow 0^+} \underline{T}_\Delta = I \text{ and } \lim_{\Delta \rightarrow 0^+} 1/\Delta(\underline{T}_\Delta - I) = \underline{Q}.$$

*Proof.* For any  $f \in \mathcal{L}(\mathcal{X})$ , it follows from Equation (5) that  $\underline{T}_\Delta f$  is continuous on  $[0, \infty)$ , which implies that  $\lim_{\Delta \rightarrow 0^+} \underline{T}_\Delta f = \underline{T}_0 f = f$ . Therefore, we infer from Proposition 1 that  $\lim_{\Delta \rightarrow 0^+} \underline{T}_\Delta = I$ , which proves the first part of this lemma. We end by proving the second part. For any  $f \in \mathcal{L}(\mathcal{X})$ , it follows from Equation (5) that

$$\lim_{\Delta \rightarrow 0^+} 1/\Delta(\underline{T}_\Delta - I)(f) = \lim_{\Delta \rightarrow 0^+} 1/\Delta(\underline{T}_\Delta f - f) = \lim_{\Delta \rightarrow 0^+} 1/\Delta(\underline{T}_\Delta f - \underline{T}_0 f) = \underline{Q}\underline{T}_0 f = \underline{Q}f.$$

Therefore, and since, for all  $\Delta > 0$ ,  $1/\Delta(\underline{T}_\Delta - I)$  is a lower transition rate operator because of Proposition 4, it follows from Proposition 5 that  $\lim_{\Delta \rightarrow 0^+} 1/\Delta(\underline{T}_\Delta - I) = \underline{Q}$ .  $\square$

*Proof of Proposition 7.* Since  $\underline{T}_0 f := f$  for all  $f \in \mathcal{L}(\mathcal{X})$ , it follows trivially that  $\underline{T}_0 = I$ . Consider now any  $t \geq 0$ . In order to prove that  $\frac{d}{dt}\underline{T}_t = \underline{Q}\underline{T}_t$ , it suffices to show that for all  $\varepsilon > 0$ , there is some  $\delta > 0$  such that

$$\left\| \frac{\underline{T}_s - \underline{T}_t}{s - t} - \underline{Q}\underline{T}_t \right\| < \varepsilon \text{ for all } s \geq 0 \text{ such that } 0 < |t - s| < \delta. \quad (11)$$

So consider any  $\varepsilon > 0$ . If  $\underline{Q} = 0$ , Equation (11) is trivially true because, since  $I$  clearly satisfies Equation (5), it follows from the unicity of the solution of Equation (5) that  $\underline{T}_t = \underline{T}_s = I$ . Therefore, in the remainder of this proof, we may assume that  $\underline{Q} \neq 0$ , which implies that  $\|\underline{Q}\| \neq 0$ . It then follows from Lemma 19 that there are  $\delta_1 > 0$  and  $\delta_2 > 0$  such that  $\|\underline{T}_q - I\| < \varepsilon/4\|\underline{Q}\|$  for all  $0 < q < \delta_1$  and  $\|1/\Delta(\underline{T}_\Delta - I) - \underline{Q}\| < \varepsilon/2$  for all  $0 < \Delta < \delta_2$ . Now define  $\delta := \min\{\delta_1, \delta_2\}$  and consider any  $s \geq 0$  such that  $0 < |t - s| < \delta$ . Let  $u := \min\{s, t\}$ ,  $\Delta := |t - s|$  and  $q := t - u$ , which implies that  $0 \leq q \leq \Delta < \delta \leq \delta_1$  and  $0 < \Delta < \delta \leq \delta_2$ . If  $q = 0$ , then  $\underline{T}_q = \underline{T}_0 = I$  and therefore  $\|\underline{Q}\underline{T}_q - \underline{Q}\| = \|\underline{Q} - \underline{Q}\| = 0$ . If  $q > 0$ , it follows from R12 and Proposition 6 that  $\|\underline{Q}\underline{T}_q - \underline{Q}\| \leq 2\|\underline{Q}\|\|\underline{T}_q - I\| < 2\|\underline{Q}\|\varepsilon/4\|\underline{Q}\| = \varepsilon/2$ . Hence, in all cases, we find that  $\|\underline{Q}\underline{T}_q - \underline{Q}\| < \varepsilon/2$ . The result now holds because

$$\begin{aligned} \left\| \frac{\underline{T}_s - \underline{T}_t}{s - t} - \underline{Q}\underline{T}_t \right\| &= \left\| \frac{\underline{T}_{\Delta+u} - \underline{T}_u}{\Delta} - \underline{Q}\underline{T}_{q+u} \right\| = \left\| \frac{\underline{T}_\Delta \underline{T}_u - \underline{T}_u}{\Delta} - \underline{Q}\underline{T}_q \underline{T}_u \right\| \\ &\leq \left\| \frac{\underline{T}_\Delta - I}{\Delta} - \underline{Q}\underline{T}_q \right\| \|\underline{T}_u\| \leq \left\| \frac{\underline{T}_\Delta - I}{\Delta} - \underline{Q}\underline{T}_q \right\| \\ &\leq \left\| \frac{\underline{T}_\Delta - I}{\Delta} - \underline{Q} \right\| + \|\underline{Q}\underline{T}_q - \underline{Q}\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where the second equality follows from Equation (6), the first inequality follows from N2 and the second inequality follows from Proposition 6 and L9.  $\square$

*Proof of Proposition 8.* The result is trivial if  $t = 0$ . In the remainder of this proof, we assume that  $t > 0$ . The result for  $\underline{Q} = 0$  is also trivial because, since  $I$  then clearly satisfies Equation (5), it follows from the unicity of the solution of Equation (5) that  $\underline{T}_t = I$ . Therefore, in the remainder of this proof, we assume that  $\underline{Q} \neq 0$ , which implies that  $\|\underline{Q}\| \neq 0$ . We will now prove that for every  $\varepsilon > 0$ , there is some  $n \in \mathbb{N}$  such that

$$\left\| \underline{T}_t - \left(I + \frac{t}{k}\underline{Q}\right)^k \right\| < \varepsilon \text{ for all } k \geq n.$$

So consider any  $\varepsilon > 0$ . It then follows from Lemma 19 that there is some  $\delta > 0$  such that  $\|1/\Delta(\underline{T}_\Delta - I) - \underline{Q}\| < \varepsilon/t$  for all  $0 < \Delta < \delta$ . Now choose  $n \in \mathbb{N}$  such that  $n > \max\{t/\delta, t\|\underline{Q}\|\}$  and consider any  $k \geq n$ . Let  $\Delta := t/k \leq t/n$ , which implies that  $0 < \Delta < \delta$  and  $\Delta\|\underline{Q}\| < 1$ . Then

$$\begin{aligned} & \left\| (\underline{T}_\Delta)^k - (I + \Delta \underline{Q})^k \right\| \\ &= \left\| (\underline{T}_\Delta)^k - (\underline{T}_\Delta)^{k-1} (I + \Delta \underline{Q}) + (\underline{T}_\Delta)^{k-1} (I + \Delta \underline{Q}) - (I + \Delta \underline{Q})^k \right\| \\ &\leq \left\| (\underline{T}_\Delta)^k - (\underline{T}_\Delta)^{k-1} (I + \Delta \underline{Q}) \right\| + \left\| (\underline{T}_\Delta)^{k-1} (I + \Delta \underline{Q}) - (I + \Delta \underline{Q})^k \right\| \\ &\leq \left\| \underline{T}_\Delta - (I + \Delta \underline{Q}) \right\| + \left\| (\underline{T}_\Delta)^{k-1} - (I + \Delta \underline{Q})^{k-1} \right\| \left\| I + \Delta \underline{Q} \right\| \\ &\leq \left\| \underline{T}_\Delta - (I + \Delta \underline{Q}) \right\| + \left\| (\underline{T}_\Delta)^{k-1} - (I + \Delta \underline{Q})^{k-1} \right\|, \end{aligned}$$

where the second inequality follows from Proposition 6 and L11 [by applying them repeatedly] and N2, and the third inequality follows from L9 and Proposition 3. By continuing in this way, we find that

$$\left\| (\underline{T}_\Delta)^k - (I + \Delta \underline{Q})^k \right\| \leq k \left\| \underline{T}_\Delta - (I + \Delta \underline{Q}) \right\|.$$

Therefore, since

$$\left\| \underline{T}_\Delta - (I + \Delta \underline{Q}) \right\| = \Delta \left\| \frac{\underline{T}_\Delta - I}{\Delta} - \underline{Q} \right\| < \Delta \frac{\varepsilon}{t} = \frac{t}{k} \frac{\varepsilon}{t} = \frac{\varepsilon}{k},$$

and because it follows from Equation (6) that  $\underline{T}_t = (\underline{T}_{t/k})^k = (\underline{T}_\Delta)^k$ , we find that

$$\left\| \underline{T}_t - (I + \frac{t}{k} \underline{Q})^k \right\| = \left\| (\underline{T}_\Delta)^k - (I + \Delta \underline{Q})^k \right\| \leq k \left\| \underline{T}_\Delta - (I + \Delta \underline{Q}) \right\| < k \frac{\varepsilon}{k} = \varepsilon.$$

□

## A.6 Proofs of results in Section 7

*Proof of Proposition 9.* First assume that  $\underline{Q}$  is ergodic. For all  $f \in \mathcal{L}(\mathcal{X})$ , it then follows from Definition 6 that  $\lim_{s \rightarrow \infty} \underline{T}_s f$  exists and is a constant function. Therefore, for all  $f \in \mathcal{L}(\mathcal{X})$ , it follows from Equation (6) that

$$\lim_{n \rightarrow \infty} \underline{T}_t^n f = \lim_{n \rightarrow \infty} \underline{T}_{nt} f = \lim_{s \rightarrow \infty} \underline{T}_s f$$

exists and is a constant function, which implies that  $\underline{T}_t$  is ergodic.

Next, assume that  $\underline{T}_t$  is ergodic. This means that, for all  $f \in \mathcal{L}(\mathcal{X})$ , there is some  $c_f \in \mathbb{R}$  such that

$$(\forall \varepsilon > 0)(\exists n \in \mathbb{N})(\forall k \geq n) \left\| \underline{T}_t^k f - c_f \right\| < \varepsilon. \quad (12)$$

Consider now any  $f \in \mathcal{L}(\mathcal{X})$  and any  $\varepsilon > 0$ . It then follows from Equation (12) that there is some  $n_\varepsilon \in \mathbb{N}$  such that  $\left\| \underline{T}_t^{n_\varepsilon} f - c_f \right\| < \varepsilon$ , which, because of Proposition 6 and L5, implies that  $\left\| \underline{T}_t^{n_\varepsilon} (f - c_f) \right\| < \varepsilon$ . Now let  $s_\varepsilon := n_\varepsilon t$ . Then for all  $s \geq s_\varepsilon$ , we have that

$$\left\| \underline{T}_s f - c_f \right\| = \left\| \underline{T}_s (f - c_f) \right\| = \left\| \underline{T}_{s-s_\varepsilon} \underline{T}_t^{n_\varepsilon} (f - c_f) \right\| \leq \left\| \underline{T}_{s-s_\varepsilon} \right\| \left\| \underline{T}_t^{n_\varepsilon} (f - c_f) \right\| < \varepsilon,$$

where the first equality follows from Proposition 6 and L5, the second equality follows from Equation (6), the first inequality follows from N1 and the last inequality follows from Proposition 6, L9 and the fact that  $\|\underline{T}_t^{n_\varepsilon} f - c_f\| < \varepsilon$ . Hence, we have found that for all  $\varepsilon > 0$ , there is some  $s_\varepsilon > 0$  such that  $\|\underline{T}_s f - c_f\| < \varepsilon$  for all  $s \geq s_\varepsilon$ . In other words:  $\lim_{s \rightarrow \infty} \underline{T}_s f = c_f$ . Since this is true for all  $f \in \mathcal{L}(\mathcal{X})$ , it follows from Definition 6 that  $\underline{Q}$  is ergodic.  $\square$

*Proof of Corollary 10.* Immediate consequence of Propositions 2, 6 and 9.  $\square$

**Lemma 20.** *Let  $\underline{Q}$  be a lower transition rate operator. Consider any  $f \in \mathcal{L}(\mathcal{X})$  and  $x \in \mathcal{X}$  such that  $f(x) > \min f$ . Then for all  $t \geq 0$ :  $\underline{T}_t f(x) > \min f$ .*

*Proof.* Since we know from Lemma 17 that  $\underline{T}_t f$  is continuously differentiable on  $[0, \infty)$ , we know that  $r_t := \underline{T}_t f - \min f$  and therefore also  $r_t(x)$  is continuously differentiable on  $[0, \infty)$ . Furthermore, for all  $t \geq 0$ , it follows from Proposition 6 and L1 that  $r_t \geq 0$ , which in turn implies that

$$\frac{d}{dt} r_t(x) = \frac{d}{dt} \underline{T}_t f(x) = \underline{Q}(\underline{T}_t f)(x) = \underline{Q}(r_t)(x) \geq \sum_{y \in \mathcal{X}} r_t(y) \underline{Q}(\mathbb{I}_y)(x) \geq r_t(x) \underline{Q}(\mathbb{I}_x)(x),$$

where the second equality follows from Equation (5), the third equality follows from R6, the first inequality follows from R2 and R3 and the last inequality follows from R4. Hence, for all  $t \geq 0$ , we find that  $r_t(x) \geq r_0(x) e^{[\underline{Q}(\mathbb{I}_x)(x)]t}$ . Since we also know that  $r_0(x) = \underline{T}_0 f(x) - \min f = f(x) - \min f > 0$ , this implies that for all  $t \geq 0$ :

$$\underline{T}_t f(x) - \min f = r_t(x) \geq r_0(x) e^{[\underline{Q}(\mathbb{I}_x)(x)]t} > 0.$$

$\square$

**Lemma 21.** *Let  $\underline{Q}$  be a lower transition rate operator. Consider any  $f \in \mathcal{L}(\mathcal{X})$ ,  $x \in \mathcal{X}$  and  $s \geq 0$  such that  $\underline{T}_s f(x) > \min f$ . Then for all  $t \geq s$ :  $\underline{T}_t f(x) > \min f$ .*

*Proof.* Because of Equation (6), it suffices to prove that  $\underline{T}_{t-s} \underline{T}_s f(x) > \min f$ . We consider two cases:  $\min \underline{T}_s f > \min f$  and  $\min \underline{T}_s f = \min f$ ;  $\min \underline{T}_s f < \min f$  is not possible because of Proposition 6 and L1. If  $\min \underline{T}_s f > \min f$ , it follows from Proposition 6 and L1 that  $\underline{T}_{t-s} \underline{T}_s f(x) \geq \min \underline{T}_s f > \min f$ . If  $\min \underline{T}_s f = \min f$ , then  $\underline{T}_s f(x) > \min \underline{T}_s f$  and therefore, because of Lemma 20,  $\underline{T}_{t-s} \underline{T}_s f(x) > \min \underline{T}_s f = \min f$ .  $\square$

**Lemma 22.** *Let  $\underline{Q}$  be a lower transition rate operator. Consider any  $f \in \mathcal{L}(\mathcal{X})$ ,  $x \in \mathcal{X}$  and  $t, s > 0$ . Then*

$$\underline{T}_t f(x) > \min f \Leftrightarrow \underline{T}_s f(x) > \min f.$$

*Proof.* For any  $\tau \geq 0$ , let

$$\mathcal{X}_\tau := \{y \in \mathcal{X} : \underline{T}_\tau f(y) > \min f\}. \quad (13)$$

It then follows from Lemma 21 that  $\mathcal{X}_\tau$  is an increasing function of  $\tau$ :

$$\tau \leq \tau' \Rightarrow \mathcal{X}_\tau \subseteq \mathcal{X}_{\tau'}. \quad (14)$$

Assume *ex absurdo* that

$$(\forall \tau' > 0) (\forall \mathcal{X}' \subseteq \mathcal{X}) (\exists \tau \in (0, \tau']) \mathcal{X}_\tau \neq \mathcal{X}'. \quad (15)$$

Choose any  $\tau_1 > 0$ . Then clearly,  $\mathcal{X}_{\tau_1} \subseteq \mathcal{X}$ . Therefore, due to Equation (15), we know that there is some  $0 < \tau_2 < \tau_1$  such that  $\mathcal{X}_{\tau_2} \neq \mathcal{X}_{\tau_1}$ , which, because of Equation (14), implies that  $\mathcal{X}_{\tau_2} \subset \mathcal{X}_{\tau_1}$ . Similarly, we infer that there is some  $0 < \tau_3 < \tau_2$  such that  $\mathcal{X}_{\tau_3} \subset \mathcal{X}_{\tau_2}$ . By continuing in this way, we obtain an infinite sequence of time points  $\tau_1 > \tau_2 > \tau_3 > \dots > \tau_i > \dots > 0$  such that  $\mathcal{X} \supseteq \mathcal{X}_{\tau_1} \supset \mathcal{X}_{\tau_2} \supset \mathcal{X}_{\tau_3} \supset \dots \supset \mathcal{X}_{\tau_i} \supset \dots$ . Since  $\mathcal{X}$  is a finite set, this is a contradiction, leading us to conclude that Equation (15) is false. This implies that there is some  $\tau^* > 0$  and  $\mathcal{X}^* \subseteq \mathcal{X}$  such that

$$(\forall \tau \in (0, \tau^*]) \mathcal{X}_\tau = \mathcal{X}^*. \quad (16)$$

Fix any  $\tau > \tau^*$  and choose  $n \in \mathbb{N}$  high enough such that  $2\tau/n \leq \tau^*$ . It then follows from Equation (16) that  $\mathcal{X}_{\tau/n} = \mathcal{X}_{2\tau/n} = \mathcal{X}^*$ . Furthermore, because of Proposition 6, L1 and Equation (13), we know that  $\underline{T}_{\tau/n}f(y) = \underline{T}_{2\tau/n}f(y) = \min f$  for all  $y \in \mathcal{X} \setminus \mathcal{X}^*$ . Therefore, we infer from Equation (13) that there is some  $\lambda > 0$  such that

$$\underline{T}_{2\tau/n}f - \min f \leq \lambda (\underline{T}_{\tau/n}f - \min f),$$

which, because of Equation (6), Proposition 6 and L5 implies that

$$\underline{T}_{\tau/n}^2(f - \min f) = \underline{T}_{2\tau/n}(f - \min f) = \underline{T}_{2\tau/n}f - \min f \leq \lambda (\underline{T}_{\tau/n}f - \min f).$$

Hence, it follows from Proposition 6, L5, Equation (6), L6 and L3 that

$$\underline{T}_\tau f - \min f = \underline{T}_\tau(f - \min f) = \underline{T}_{\tau/n}^n(f - \min f) \leq \lambda^{n-1}(\underline{T}_{\tau/n}f - \min f). \quad (17)$$

Consider now any  $y \in \mathcal{X} \setminus \mathcal{X}^*$ . Since  $\underline{T}_{\tau/n}f(y) = \min f$ , it follows from Equation (17) that  $\underline{T}_\tau f(y) \leq \min f$ , which in turn implies that  $y \notin \mathcal{X}_\tau$ . Since this holds for all  $y \in \mathcal{X} \setminus \mathcal{X}^*$ , we find that  $\mathcal{X}_\tau \subseteq \mathcal{X}^* = \mathcal{X}_{\tau/n}$ . Furthermore, since  $\tau/n \leq \tau$ , it follows from Equation (14) that  $\mathcal{X}_{\tau/n} \subseteq \mathcal{X}_\tau$ . Hence, we find that  $\mathcal{X}_\tau = \mathcal{X}^*$ . Since this is true for all  $\tau > \tau^*$ , it follows from Equation (16) that

$$\mathcal{X}_\tau = \mathcal{X}^* \text{ for all } \tau > 0.$$

Therefore, due to Equation (13), we find that

$$\underline{T}_t f(x) > \min f \Leftrightarrow x \in \mathcal{X}_t \Leftrightarrow x \in \mathcal{X}_s \Leftrightarrow \underline{T}_s f(x) > \min f.$$

□

**Lemma 23.** *Let  $\underline{Q}$  be a lower transition rate operator. Consider any  $f \in \mathcal{L}(\mathcal{X})$ ,  $x \in \mathcal{X}$  and  $s \geq 0$  such that  $\overline{T}_s f(x) > \min f$ . Then for all  $t \geq s$ :  $\overline{T}_t f(x) > \min f$ .*

*Proof.* Because of Equations (3) and (6), it suffices to prove that  $\overline{T}_{t-s} \overline{T}_s f(x) > \min f$ . We consider two cases:  $\min \overline{T}_s f > \min f$  and  $\min \overline{T}_s f = \min f$ ;  $\min \overline{T}_s f < \min f$  is not possible because of Proposition 6 and L4. If  $\min \overline{T}_s f > \min f$ , it follows from Proposition 6 and L4 that  $\overline{T}_{t-s} \overline{T}_s f(x) \geq \min \overline{T}_s f > \min f$ . If  $\min \overline{T}_s f = \min f$ , then  $\overline{T}_s f(x) > \min \overline{T}_s f$  and therefore, it follows from Proposition 6, L4 and Lemma 20 that  $\overline{T}_{t-s} \overline{T}_s f(x) \geq \underline{T}_{t-s} \overline{T}_s f(x) > \min \overline{T}_s f = \min f$ . □

**Lemma 24.** *Let  $\underline{Q}$  be a lower transition rate operator. Consider any  $f \in \mathcal{L}(\mathcal{X})$ ,  $x \in \mathcal{X}$  and  $t, s > 0$ . Then*

$$\overline{T}_t f(x) > \min f \Leftrightarrow \overline{T}_s f(x) > \min f.$$

*Proof.* For any  $\tau \geq 0$ , let

$$\mathcal{X}_\tau := \{y \in \mathcal{X} : \overline{T}_\tau f(y) > \min f\}. \quad (18)$$

It then follows from Lemma 23 that  $\mathcal{X}_\tau$  is an increasing function of  $\tau$ :

$$\tau \leq \tau' \Rightarrow \mathcal{X}_\tau \subseteq \mathcal{X}_{\tau'}. \quad (19)$$

Using an argument that is identical to that in Lemma 22, we find that this implies that there is some  $\tau^* > 0$  and  $\mathcal{X}^* \subseteq \mathcal{X}$  such that

$$(\forall \tau \in (0, \tau^*]) \mathcal{X}_\tau = \mathcal{X}^*. \quad (20)$$

Fix any  $\tau > \tau^*$  and choose  $n \in \mathbb{N}$  high enough such that  $2\tau/n \leq \tau^*$ . It then follows from Equation (20) that  $\mathcal{X}_{\tau/n} = \mathcal{X}_{2\tau/n} = \mathcal{X}^*$ . Furthermore, because of Proposition 6, L4 and Equation (18), we know that  $\overline{T}_{\tau/n} f(y) = \overline{T}_{2\tau/n} f(y) = \min f$  for all  $y \in \mathcal{X} \setminus \mathcal{X}^*$ . Therefore, we infer from Equation (18) that there is some  $\lambda > 0$  such that

$$\overline{T}_{2\tau/n} f - \min f \leq \lambda (\overline{T}_{\tau/n} f - \min f),$$

which, because of Equations (3) and (6), Proposition 6 and L5 implies that

$$\overline{T}_{\tau/n}^2 (f - \min f) = \overline{T}_{2\tau/n} (f - \min f) = \overline{T}_{2\tau/n} f - \min f \leq \lambda (\overline{T}_{\tau/n} f - \min f).$$

Hence, it follows from Proposition 6, L5, Equations (3) and (6), L6 and L3 that

$$\overline{T}_\tau f - \min f = \overline{T}_\tau (f - \min f) = \overline{T}_{\tau/n}^n (f - \min f) \leq \lambda^{n-1} (\overline{T}_{\tau/n} f - \min f). \quad (21)$$

Consider now any  $y \in \mathcal{X} \setminus \mathcal{X}^*$ . Since  $\overline{T}_{\tau/n} f(y) = \min f$ , it follows from Equation (21) that  $\overline{T}_\tau f(y) \leq \min f$ , which in turn implies that  $y \notin \mathcal{X}_\tau$ . Since this holds for all  $y \in \mathcal{X} \setminus \mathcal{X}^*$ , we find that  $\mathcal{X}_\tau \subseteq \mathcal{X}^* = \mathcal{X}_{\tau/n}$ . Furthermore, since  $\tau/n \leq \tau$ , it follows from Equation (19) that  $\mathcal{X}_{\tau/n} \subseteq \mathcal{X}_\tau$ . Hence, we find that  $\mathcal{X}_\tau = \mathcal{X}^*$ . Since this is true for all  $\tau > \tau^*$ , it follows from Equation (20) that

$$\mathcal{X}_\tau = \mathcal{X}^* \text{ for all } \tau > 0.$$

Therefore, due to Equation (18), we find that

$$\overline{T}_t f(x) > \min f \Leftrightarrow x \in \mathcal{X}_t \Leftrightarrow x \in \mathcal{X}_s \Leftrightarrow \overline{T}_s f(x) > \min f.$$

□

**Proposition 25.** Let  $\underline{Q}$  be a lower transition rate operator. Then for all  $f \in \mathcal{L}(\mathcal{X})$ ,  $x \in \mathcal{X}$  and  $t, s > 0$ :

$$\begin{aligned} f(x) > \min f &\Rightarrow \underline{T}_t f(x) > \min f \Leftrightarrow \underline{T}_s f(x) > \min f; \\ f(x) < \max f &\Rightarrow \underline{T}_t f(x) < \max f \Leftrightarrow \underline{T}_s f(x) < \max f; \\ f(x) > \min f &\Rightarrow \overline{T}_t f(x) > \min f \Leftrightarrow \overline{T}_s f(x) > \min f; \\ f(x) < \max f &\Rightarrow \overline{T}_t f(x) < \max f \Leftrightarrow \overline{T}_s f(x) < \max f; \end{aligned}$$

*Proof.* The first implication  $[f(x) > \min f \Rightarrow \underline{T}_t f(x) > \min f]$  follows from Lemma 20 and the first equivalence  $[\underline{T}_t f(x) > \min f \Leftrightarrow \underline{T}_s f(x) > \min f]$  follows from Lemma 22. Since  $\overline{T}_0 f = f$ , the third implication  $[f(x) > \min f \Rightarrow \overline{T}_t f(x)]$  follows from Lemma 23. The third equivalence  $[\overline{T}_t f(x) > \min f \Leftrightarrow \overline{T}_s f(x) > \min f]$  follows from Lemma 24. The rest of the result now follows directly because we know from Equation (3) that  $\overline{T}_t f(x) = -\underline{T}_t(-f)(x)$ ,  $\overline{T}_s f(x) = -\underline{T}_s(-f)(x)$  and  $\max f = -\min(-f)$ . □

**Corollary 26.** Let  $\underline{Q}$  be a lower transition rate operator. Then for all  $A \subseteq \mathcal{X}$ , all  $x \in \mathcal{X}$  and all  $t, s > 0$ :

$$\begin{aligned} x \in A &\Rightarrow \underline{T}_t \mathbb{I}_A(x) > 0 \Leftrightarrow \underline{T}_s \mathbb{I}_A(x) > 0; \\ x \notin A &\Rightarrow \underline{T}_t \mathbb{I}_A(x) < 1 \Leftrightarrow \underline{T}_s \mathbb{I}_A(x) < 1; \\ x \in A &\Rightarrow \overline{T}_t \mathbb{I}_A(x) > 0 \Leftrightarrow \overline{T}_s \mathbb{I}_A(x) > 0; \\ x \notin A &\Rightarrow \overline{T}_t \mathbb{I}_A(x) < 1 \Leftrightarrow \overline{T}_s \mathbb{I}_A(x) < 1. \end{aligned}$$

*Proof.* If  $A = \emptyset$  or  $A = \mathcal{X}$ , the result follows trivially from Proposition 6 and L4. In all other cases, the result follows directly from Proposition 25, with  $f = \mathbb{I}_A$ .  $\square$

*Proof of Proposition 11.* If  $t = 0$ , we know from Proposition 7 that  $\underline{T}_t = I$ , which implies that  $\underline{T}_t^n = I = \underline{T}_t$  and  $\overline{T}_t^n = I = \overline{T}_t$  for all  $n \in \mathbb{N}$ . In that case, Definitions 3 and 4 are trivially equal. If  $t > 0$ , then since we know from Equations (6) and (3) that  $\underline{T}_t^n = \underline{T}_{nt}$  and  $\overline{T}_t^n = \overline{T}_{nt}$  for all  $n \in \mathbb{N}$ , the equivalence of Definitions 3 and 4 follows directly from Corollary 26.  $\square$

*Proof of Corollary 12.* This result is a trivial consequence of Corollary 10 and Proposition 11.  $\square$

*Proof of Proposition 13.* First assume that  $\overline{T}_t \mathbb{I}_x(y) > 0$ . It then follows from Proposition 8 and Equations (3) and (4) that there is some  $n \in \mathbb{N}$  such that  $n \geq t \|\underline{Q}\|$  and

$$\left( \left( I + \frac{t}{n} \overline{Q} \right)^n \mathbb{I}_x \right)(y) > 0. \quad (22)$$

Let  $\Delta := t/n \geq 0$  and define  $\underline{T}_* := I + \Delta \underline{Q}$ . Since  $n \geq t \|\underline{Q}\|$  implies that  $\Delta \|\underline{Q}\| \leq 1$ , it then follows from Proposition 3 that  $\underline{T}_*$  is a lower transition operator. Therefore, for all  $z \in \mathcal{X}$  and  $w \in \mathcal{X}$ , it follows from L4 that  $c(w, z) := (\overline{T}_* \mathbb{I}_z)(w) \geq 0$ . For all  $z \in \mathcal{X}$ , we now have that

$$\overline{T}_* \mathbb{I}_z = \sum_{w \in \mathcal{X}} \mathbb{I}_w \cdot (\overline{T}_* \mathbb{I}_z)(w) = \sum_{w \in \mathcal{X}} c(w, z) \mathbb{I}_w.$$

Hence, for all  $x_n \in \mathcal{X}$ , it follows from Equation (3), L2 and L6 that

$$\overline{T}_*^n \mathbb{I}_{x_n} = \overline{T}_*^{n-1} \overline{T}_* \mathbb{I}_{x_n} = \overline{T}_*^{n-1} \sum_{x_{n-1} \in \mathcal{X}} c(x_{n-1}, x_n) \mathbb{I}_{x_{n-1}} \leq \sum_{x_{n-1} \in \mathcal{X}} c(x_{n-1}, x_n) \overline{T}_*^{n-1} \mathbb{I}_{x_{n-1}}$$

and, by continuing in this way, that

$$\overline{T}_*^n \mathbb{I}_{x_n} \leq \sum_{x_{n-1} \in \mathcal{X}} c(x_{n-1}, x_n) \sum_{x_{n-2} \in \mathcal{X}} c(x_{n-2}, x_{n-1}) \cdots \sum_{x_1 \in \mathcal{X}} c(x_1, x_2) \overline{T}_* \mathbb{I}_{x_1}.$$

Therefore, for all  $x_n \in \mathcal{X}$  and  $x_0 \in \mathcal{X}$ , we find that

$$(\overline{T}_*^n \mathbb{I}_{x_n})(x_0) \leq \sum_{x_{n-1} \in \mathcal{X}} c(x_{n-1}, x_n) \sum_{x_{n-2} \in \mathcal{X}} c(x_{n-2}, x_{n-1}) \cdots \sum_{x_1 \in \mathcal{X}} c(x_1, x_2) c(x_0, x_1).$$

Hence, if we let  $x_0 := y$  and  $x_n := x$ , it follows from Equation (22) that

$$\sum_{x_{n-1} \in \mathcal{X}} c(x_{n-1}, x_n) \sum_{x_{n-2} \in \mathcal{X}} c(x_{n-2}, x_{n-1}) \cdots \sum_{x_1 \in \mathcal{X}} c(x_1, x_2) c(x_0, x_1) > 0.$$



This implies that there is some sequence  $y = x_0, x_1, \dots, x_n = x$  such that

$$c(x_{n-1}, x_n) c(x_{n-2}, x_{n-1}) \cdots c(x_1, x_2) c(x_0, x_1) > 0.$$

Since each of the factors in this product is non-negative, it follows that  $c(x_{k-1}, x_k) > 0$  for all  $k \in \{1, \dots, n\}$ . Therefore, for any  $k \in \{1, \dots, n\}$  such that  $x_k \neq x_{k-1}$ , it follows that

$$\begin{aligned} \overline{Q}(\mathbb{I}_{x_k})(x_{k-1}) &= \frac{1}{\Delta} \left( \mathbb{I}_{x_k}(x_{k-1}) + \Delta \overline{Q}(\mathbb{I}_{x_k})(x_{k-1}) \right) \\ &= \frac{1}{\Delta} \left( (\mathbb{I}_{x_k} + \Delta \overline{Q}(\mathbb{I}_{x_k}))(x_{k-1}) \right) \\ &= \frac{1}{\Delta} \left( (I + \Delta \overline{Q}) \mathbb{I}_{x_k}(x_{k-1}) \right) = \frac{1}{\Delta} \left( (\overline{T}_* \mathbb{I}_{x_k})(x_{k-1}) \right) = \frac{1}{\Delta} c(x_{k-1}, x_k) > 0. \end{aligned}$$

If  $x_k \neq x_{k-1}$  for all  $k \in \{1, \dots, n\}$ , this implies that  $x$  is upper reachable from  $y$ . Otherwise, let  $x'_0, \dots, x'_m$  be a new sequence, obtained by removing from  $x_0, \dots, x_n$  those elements  $x_k$  for which  $x_k = x_{k-1}$ ;  $n - m$  is the number of elements that is removed. Then  $x'_0 = y$ ,  $x'_m = x$  and, for all  $k \in \{1, \dots, m\}$ , we have that  $x'_k \neq x'_{k-1}$  and  $\overline{Q}(\mathbb{I}_{x'_k})(x'_{k-1}) > 0$ . Therefore,  $x$  is upper reachable from  $y$ .

Conversely, assume that  $x$  is upper reachable from  $y$ , meaning that there is some sequence  $y = x_0, x_1, \dots, x_n = x$  such that, for all  $k \in \{1, \dots, n\}$ ,  $x_k \neq x_{k-1}$  and  $\overline{Q}(\mathbb{I}_{x_k})(x_{k-1}) > 0$ . If  $n = 0$ , then  $x = y$  and therefore, it follows from Corollary 26 that  $\overline{T}_t \mathbb{I}_x(y) > 0$ . Hence, for the remainder of this proof, we may assume that  $n \geq 1$ . Fix any  $k \in \{1, \dots, n\}$ . We then have that  $\overline{T}_0 \mathbb{I}_{x_k}(x_{k-1}) = \mathbb{I}_{x_k}(x_{k-1}) = 0$  and

$$\begin{aligned} \left. \frac{d}{ds} \overline{T}_s \mathbb{I}_{x_k}(x_{k-1}) \right|_{s=0} &= - \left. \frac{d}{ds} \mathcal{T}_s(-\mathbb{I}_{x_k})(x_{k-1}) \right|_{s=0} \\ &= - \underline{Q}(\mathcal{T}_0(-\mathbb{I}_{x_k}))(x_{k-1}) = - \underline{Q}(-\mathbb{I}_{x_k})(x_{k-1}) = \overline{Q}(\mathbb{I}_{x_k})(x_{k-1}) > 0, \end{aligned}$$

where the first equality follows from Equation (3), the second equality follows from Equation (5) and the last equality follows from Equation (4). Therefore, there is some  $\varepsilon_k > 0$  such that  $\overline{T}_{\varepsilon_k} \mathbb{I}_{x_k}(x_{k-1}) > 0$ . Consequently, if we let  $c_k := \overline{T}_{\varepsilon_k} \mathbb{I}_{x_k}(x_{k-1}) > 0$ , then because it follows from Proposition 6 and L4 that  $\overline{T}_{\varepsilon_k} \mathbb{I}_{x_k} \geq 0$ , we have that  $\overline{T}_{\varepsilon_k} \mathbb{I}_{x_k} \geq c_k \mathbb{I}_{x_{k-1}}$ . Let  $\varepsilon := \sum_{k=1}^n \varepsilon_k > 0$ . Then

$$\overline{T}_\varepsilon \mathbb{I}_{x_n} = \overline{T}_{\varepsilon_1} \cdots \overline{T}_{\varepsilon_{n-1}} \overline{T}_{\varepsilon_n} \mathbb{I}_{x_n} \geq c_n \overline{T}_{\varepsilon_1} \cdots \overline{T}_{\varepsilon_{n-1}} \mathbb{I}_{x_{n-1}} \geq \cdots \geq \left( \prod_{k=1}^n c_k \right) \mathbb{I}_{x_0},$$

where the equality follows from Equations (3) and (6) and where the inequalities follow from Proposition 6, L6, L3 and Equation (3). Therefore, we find that

$$\overline{T}_\varepsilon \mathbb{I}_x(y) = \overline{T}_\varepsilon \mathbb{I}_{x_n}(x_0) \geq \prod_{k=1}^n c_k \mathbb{I}_{x_0}(x_0) = \prod_{k=1}^n c_k > 0,$$

which implies that  $\overline{T}_t \mathbb{I}_x(y) > 0$  because of Corollary 26.  $\square$

*Proof of Proposition 14.* Let  $\{A_k\}_{k \in \mathbb{N}_0}$  and  $n$  be defined as in Definition 8. We need to prove that  $x \in A_n$  if and only if  $\underline{T}_t \mathbb{I}_A(x) > 0$ .

First assume that  $\underline{T}_t \mathbb{I}_A(x) > 0$ . It then follows from Proposition 8 that there is some  $m \in \mathbb{N}$  such that  $m \geq n$ ,  $m \geq t \|\underline{Q}\|$  and

$$\left( \left( I + \frac{t}{m} \underline{Q} \right)^m \mathbb{I}_A \right)(x) > 0. \quad (23)$$

Let  $\Delta := t/m \geq 0$  and define  $\underline{T}_* := I + \Delta \underline{Q}$ . Since  $m \geq t \|\underline{Q}\|$  implies that  $\Delta \|\underline{Q}\| \leq 1$ , it then follows from Proposition 3 that  $\underline{T}_*$  is a lower transition operator. Consider any  $k \in \mathbb{N}_0$  such that  $k \leq m$  and any  $y \in \mathcal{X} \setminus A_{k+1}$ . Since  $A_k \subseteq A_{k+1}$ , this implies that  $y \notin A_k$ . Assume *ex absurdo* that  $\underline{Q}(\mathbb{I}_{A_k})(y) > 0$ . It then follows from Equation (8) that  $y \in A_{k+1}$ , a contradiction. Hence, we find that  $\underline{Q}(\mathbb{I}_{A_k})(y) \leq 0$ . Since  $y \notin A_k$ , it follows from R2 and R4 that  $\underline{Q}(\mathbb{I}_{A_k})(y) \geq \sum_{z \in A_k} \underline{Q}(\mathbb{I}_z)(y) \geq 0$ . Hence, we infer that  $\underline{Q}(\mathbb{I}_{A_k})(y) = 0$ . Furthermore, since  $y \notin A_k$ , we also have that  $\mathbb{I}_{A_k}(y) = 0$ . Hence, we find that  $(\underline{T}_* \mathbb{I}_{A_k})(y) = ((I + \Delta \underline{Q}) \mathbb{I}_{A_k})(y) = 0$ . Since this holds for all  $y \in \mathcal{X} \setminus A_{k+1}$ , there is some  $c_k > 0$  such that  $\underline{T}_* \mathbb{I}_{A_k} \leq c_k \mathbb{I}_{A_{k+1}}$ . Due to L3 and L6, this implies that  $\underline{T}_*^{m-k} \mathbb{I}_{A_k} \leq c_k \underline{T}_*^{m-k-1} \mathbb{I}_{A_{k+1}}$ . Since this holds for all  $k \in \mathbb{N}_0$  such that  $k \leq m$ , we find that

$$\underline{T}_*^m \mathbb{I}_{A_0} \leq c_0 \underline{T}_*^{m-1} \mathbb{I}_{A_1} \leq c_0 c_1 \underline{T}_*^{m-2} \mathbb{I}_{A_2} \leq \dots \leq c_0 c_1 \dots c_{m-1} \mathbb{I}_{A_m}.$$

Therefore, since  $A_0 = A$ , it follows from Equation (23) that  $x \in A_m$ . Since  $A_n = A_{n+1}$ , it follows from Equation (8) that  $A_r = A_n$  for all  $r \geq n$  and therefore, in particular, that  $A_m = A_n$ . Since  $x \in A_m$ , this implies that  $x \in A_n$ .

Conversely, assume that  $x \in A_n$ . If  $n = 0$ , then  $A_n = A_0 = A$  and therefore  $x \in A$ , which, due to Corollary 26, implies that  $\underline{T}_* \mathbb{I}_A(x) > 0$ . Therefore, for the remainder of this proof, we may assume that  $n \geq 1$ . Fix any  $k \in \{0, \dots, n-1\}$ . Consider any  $y \in A_{k+1} \setminus A_k$ . Then  $\underline{T}_0 \mathbb{I}_{A_k}(y) = \mathbb{I}_{A_k}(y) = 0$  and

$$\left. \frac{d}{ds} \underline{T}_s \mathbb{I}_{A_k}(y) \right|_{s=0} = \underline{Q}(\underline{T}_0(\mathbb{I}_{A_k}))(y) = \underline{Q}(\mathbb{I}_{A_k})(y) > 0,$$

where the first equality follows from Equation (5) and the inequality follows from Equation (8). Therefore, there is some  $\varepsilon_{k,y} > 0$  such that  $\underline{T}_s \mathbb{I}_{A_k}(y) > 0$  for all  $s \in (0, \varepsilon_{k,y}]$ . Hence, if we let  $\varepsilon_k := \min_{y \in A_{k+1} \setminus A_k} \varepsilon_{k,y}$ , then  $\underline{T}_{\varepsilon_k} \mathbb{I}_{A_k}(y) > 0$  for all  $y \in A_{k+1} \setminus A_k$ . For all  $y \in A_k$ , it follows from Corollary 26 that  $\underline{T}_{\varepsilon_k} \mathbb{I}_{A_k}(y) > 0$ . Hence, in summary, we have that  $\underline{T}_{\varepsilon_k} \mathbb{I}_{A_k}(y) > 0$  for all  $y \in A_{k+1}$ . Since we know from Proposition 6 and L1 that  $\underline{T}_{\varepsilon_k} \mathbb{I}_{A_k} \geq 0$ , this implies that there is some  $c_k > 0$  such that  $\underline{T}_{\varepsilon_k} \mathbb{I}_{A_k} \geq c_k \mathbb{I}_{A_{k+1}}$ . Let  $\varepsilon := \sum_{k=0}^{n-1} \varepsilon_k$ . Then

$$\underline{T}_\varepsilon \mathbb{I}_{A_0} = \underline{T}_{\varepsilon_{n-1}} \dots \underline{T}_{\varepsilon_1} \underline{T}_{\varepsilon_0} \mathbb{I}_{A_0} \geq c_0 \underline{T}_{\varepsilon_{n-1}} \dots \underline{T}_{\varepsilon_1} \mathbb{I}_{A_1} \geq \dots \geq \left( \prod_{k=0}^{n-1} c_k \right) \mathbb{I}_{A_n},$$

where the equality follows from Equation (6) and the inequalities follow from Proposition 6, L3 and L6. Therefore, we find that

$$\underline{T}_\varepsilon \mathbb{I}_A(x) = \underline{T}_\varepsilon \mathbb{I}_{A_0}(x) \geq \left( \prod_{k=0}^{n-1} c_k \right) \mathbb{I}_{A_n}(x) = \prod_{k=0}^{n-1} c_k > 0,$$

which implies that  $\underline{T}_* \mathbb{I}_A(x) > 0$  because of Corollary 26.  $\square$

*Proof of Theorem 15.* Fix any  $t > 0$ . It then follows from Corollary 12 that  $\underline{Q}$  is ergodic if and only if  $\underline{T}_t$  is 1-step absorbing or, equivalently, because of Definition 4, if

$$\mathcal{X}_{1A} := \{x \in \mathcal{X} : \min \overline{T}_t \mathbb{I}_x > 0\} \neq \emptyset \text{ and } (\forall x \in \mathcal{X} \setminus \mathcal{X}_{1A}) \underline{T}_t \mathbb{I}_{\mathcal{X}_{1A}}(x) > 0.$$

The result now follows immediately because we know from Proposition 13 that, for all  $x \in \mathcal{X}$ ,

$$\min \overline{T}_t \mathbb{I}_x > 0 \Leftrightarrow (\forall y \in \mathcal{X}) \overline{T}_t \mathbb{I}_x(y) > 0 \Leftrightarrow (\forall y \in \mathcal{X}) y \dot{\rightarrow} x,$$

and because we know from Proposition 14 that, for all  $x \in \mathcal{X} \setminus \mathcal{X}_{1A}$ ,  $\underline{T}_t \mathbb{I}_{\mathcal{X}_{1A}}(x) > 0$  if and only if  $x \dot{\rightarrow} \mathcal{X}_{1A}$ .  $\square$